



TESIS - SS14 2501

***MODEL-BASED CLUSTERING DENGAN DISTRIBUSI  
 $t$  MULTIVARIAT MENGGUNAKAN KRITERIA  
INTEGRATED COMPLETED LIKELIHOOD DAN  
MINIMUM MESSAGE LENGTH***  
(Pengelompokan Provinsi di Indonesia Menurut  
Indikator Pasar Tenaga Kerja Tahun 2012-2015)

METY AGUSTINI  
NRP. 1315 201 717

DOSEN PEMBIMBING  
Dr. Kartika Fithriasari, M.Si.  
Dr. rer.pol. Dedy Dwi Prastyo, M.Si.

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017



TESIS - SS142501

***MODEL-BASED CLUSTERING DENGAN DISTRIBUSI  
 $t$  MULTIVARIAT MENGGUNAKAN KRITERIA  
INTEGRATED COMPLETED LIKELIHOOD DAN  
MINIMUM MESSAGE LENGTH***

(Pengelompokan Provinsi di Indonesia Menurut  
Indikator Pasar Tenaga Kerja Tahun 2012-2015)

METY AGUSTINI  
NRP. 1315 201 717

DOSEN PEMBIMBING  
Dr. Kartika Fithriasari, M.Si.  
Dr. rer.pol. Dedy Dwi Prastyo, M.Si.

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017



THESIS - SS142501

**MODEL- BASED CLUSTERING WITH  
THE MULTIVARIATE  $t$  DISTRIBUTION USING  
INTEGRATED COMPLETED LIKELIHOOD AND  
MINIMUM MESSAGE LENGTH CRITERIA**

(Grouping Provinces in Indonesia According to The Key  
Indicators of The Labor Market in 2012-2015)

METY AGUSTINI  
NRP. 1315 201 717

SUPERVISORS  
Dr. Kartika Fithriasari, M.Si.  
Dr. rer.pol. Dedy Dwi Prastyo, M.Si.

PROGRAM OF MAGISTER  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017

**MODEL-BASED CLUSTERING DENGAN  
DISTRIBUSI / MULTIVARIAT MENGGUNAKAN  
KRITERIA INTEGRATED COMPLETED LIKELIHOOD DAN  
MINIMUM MESSAGE LENGTH**

**(Pengelompokan Provinsi di Indonesia Menurut Indikator  
Pasar Tenaga Kerja Tahun 2012-2015)**

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Sains (M.Si)  
di  
Institut Teknologi Sepuluh Nopember  
Oleh:

**METY AGUSTINI  
NRP. 1315 201 717**

Tanggal Ujian : 12 Januari 2017  
Periode Wisuda : Maret 2017

Disetujui Oleh :

1. Dr. Kartika Fithriasari, M.Si  
NIP. 19691212 199303 2 002

(Pembimbing I)

2. Dr. rer. pol. Dedy Dwi Prastyo, M.Si  
NIP. 19831204 200812 1 002

(Pembimbing II)

3. Dr. Irfanah, M.Si  
NIP. 19780406 200112 2 002

(Penguji)

4. Dr. Vita Ratnasari, M.Si  
NIP. 19700910 199702 2 001

(Penguji)

5. Dr. Margaretha Ari Anggorowati, S.Kom, MT.  
NIP. 19720222 199803 2 002

(Penguji)

Direktur Program Pascasarjana

an. Direktur Program Pascasarjana  
Asisten Direktur



Prof. Dr. Ir. Tri Widjaja, M.Eng.  
NIP. 19611121 198603 1 001

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.  
NIP 19601202 198701 1 001

**MODEL-BASED CLUSTERING DENGAN DISTRIBUSI  
 $t$  MULTIVARIAT MENGGUNAKAN KRITERIA  
INTEGRATED COMPLETED LIKELIHOOD DAN  
MINIMUM MESSAGE LENGTH  
(Pengelompokan Provinsi di Indonesia Menurut Indikator Pasar  
Tenaga Kerja Tahun 2012-2015)**

Nama Mahasiswa	: Mety Agustini
NRP	: 1315201717
Dosen Pembimbing I	: Dr. Kartika Fithriasari
Dosen Pembimbing II	: Dr. rer.pol. Dedy Dwi Prastyo

**ABSTRAK**

Analisis *cluster* merupakan alat statistik yang banyak digunakan untuk menentukan kelompok dalam satu kumpulan data. Metode *clustering* yang paling sering digunakan adalah *clustering* berdasarkan ukuran jarak. Namun pengelompokan menggunakan jarak akan sangat sulit dilakukan jika kondisi objek yang ada saling tumpang tindih. Penelitian ini menyarankan pendekatan *model-based clustering* (MBC) yang didasarkan pada model *finite mixture*. Metode *clustering* ini memiliki asumsi bahwa data yang dihasilkan berasal dari beberapa distribusi probabilitas dan kemudian kelompok yang terbentuk diwakili oleh masing-masing distribusi probabilitas tersebut. Distribusi  $t$  multivariat pada *model-based clustering* digunakan untuk mengakomodasi keberadaan *outlier*. Distribusi  $t$  multivariat dianggap lebih tepat mengatasi *outlier* dibandingkan distribusi normal multivariat. Pemilihan model terbaik dari beberapa model yang tersedia dilakukan melalui kriteria *Integrated Completed Likelihood* (ICL) dan *Minimum Message Length* (MML). Kelompok optimal MBC-ICL digunakan untuk analisis pasar tenaga kerja Indonesia berdasarkan indikator Bekerja Menurut Lapangan Usaha. Sedangkan kelompok optimal RMBC-MML digunakan pada analisis pasar tenaga kerja Indonesia berdasarkan indikator Rasio Penduduk Bekerja Terhadap Jumlah Penduduk (EPR), Pekerja Rentan, dan Pekerja Sektor Informal.

**Kata kunci** : distribusi  $t$  multivariat, indikator pasar tenaga kerja, *integrated completed likelihood*, *minimum message length*, *model-based clustering*

**MODEL-BASED CLUSTERING  
WITH THE MULTIVARIATE  $t$  DISTRIBUTION USING  
INTEGRATED COMPLETED LIKELIHOOD AND  
MINIMUM MESSAGE LENGTH CRITERIA  
(Grouping Provinces in Indonesia According To The Key  
Indicators of The Labor Market in 2012-2015)**

Name	: Mety Agustini
Registration Number	: 1315201717
Supervisor	: Dr. Kartika Fithriasari
Co-Supervisor	: Dr. rer.pol. Dedy Dwi Prastyo

**ABSTRACT**

The cluster analysis is a widely used statistical tool to determine subsets in a given data set. Clustering methods are used mostly based on distance measures. However, the measurement by the distance will be very difficult to do if the objects overlap. This research reviews recently suggested approaches to model-based clustering (MBC) which based on finite mixture models. It has an assumption that data are generated from several probability distributions and then a different cluster is represented by each probability distribution. The multivariate  $t$  distribution in a model-based clustering is used to accomodate the existence of outlier. It is considered more appropriately overcoming the outlier than multivariate normal distribution. The best model from a list of candidate models is determined by the model selection approach : integrated completed likelihood (ICL) and minimum message length (MML) criteria. Optimal cluster of MBC-ICL is used to review the analysis of Indonesia's labor market based on indicators of employment by sector. While the optimal cluster of RMBC-MML is used to analyse the Indonesia's labor market based on indicators of The Employment to Population Ratio (EPR), The Vulnerable Employment, and The Employment In The Informal Economy.

**Keywords** : multivariate  $t$  distribution, the key indicators of the labor market, integrated completed likelihood, minimum message length, model-based clustering,

## KATA PENGANTAR

Alhamdulillah Robbil ‘Aalamiin, puji syukur kehadiran Allah SWT, atas berkat dan hidayah-Nya kepada Penulis sehingga tesis yang berjudul “*MODEL-BASED CLUSTERING DENGAN DISTRIBUSI  $t$  MULTIVARIAT MENGGUNAKAN KRITERIA INTEGRATED COMPLETED LIKELIHOOD DAN MINIMUM MESSAGE LENGTH* (Pengelompokan Provinsi di Indonesia Menurut Indikator Pasar Tenaga Kerja Tahun 2012-2015)” dapat terselesaikan. Penyelesaian tesis ini tidak lepas dari bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu, teriring rasa syukur dan doa, Penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Kartika Fithriasari, M.Si., dan Bapak Dr. rer.pol Dedy Dwi Prastyo, M.Si., selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, arahan, pembelajaran dan koreksi dengan penuh kesabaran selama penyusunan tesis ini.
2. Ibu Dr. Irhamah, M.Si., Ibu Dr. Vita Ratnasari, M.Si., dan Ibu Dr. Margaretha Ari Anggorowati, S.Kom, MT., selaku dosen penguji yang memberikan koreksi, saran dan masukan dalam penyusunan tesis ini.
3. Bapak Dr. Suhartono, M.Sc., selaku Ketua Jurusan Statistika, Bapak Dr.rer.pol. Heri Kuswanto, M.Si., selaku Ketua Program Studi Pascasarjana Jurusan Statistika, Bapak Dr. Muhammad Mashuri, MT., selaku dosen wali selama Penulis menuntut ilmu, seluruh Bapak/Ibu dosen pengajar yang telah memberikan ilmu pengetahuan dan pengalaman yang bermanfaat, serta segenap karyawan keluarga besar Jurusan Statistika FMIPA ITS Surabaya, atas segala dukungan dan bantuan yang diberikan selama Penulis menjadi bagian dari sistem.
4. Kepala BPS RI beserta jajarannya, Kepala Pusdiklat BPS beserta jajarannya, Kepala BPS Provinsi Kepulauan Bangka Belitung beserta jajarannya, Kepala BPS Kabupaten Bangka beserta jajarannya, yang telah memberikan kesempatan dan kepercayaan kepada Penulis sehingga dapat melanjutkan studi program S2 di ITS.



5. Kedua orang tua, Alm. Ayah yang sangat Penulis kasihi dan Umak yang sangat Penulis sayangi, yang telah membesarkan, mendidik dan selalu mendo'akan penulis dengan penuh keikhlasan. Penuh syukur Penulis panjatkan kepada Allah SWT untuk semua cinta dan kasih sayang kalian.
6. Muhammad Kasiyyanto Assahid, suami, sahabat dan rekan meraih cita-cita kebaikan di dunia dan akhirat. Terima kasih atas izin, kepercayaan dan ridhonya. Anak-anakku tersayang Latifah Az-zahra, Muhammad Fadhlán Al Farizi, dan Abdurrahman Al Ghifari. Kalian adalah motivator terbesar yang selalu menjadi penguat.
7. Kak Sulis sayang. Terima kasih untuk semua cinta kepada anak-anak. Semoga berbuah pahala. Kak Susi, Kak Teri, Kak Tini, Busu Tata dan keluarga atas motivasi dan do'a dalam proses penyelesaian tesis ini.
8. Sahabat sahabat di ARH48: Mbak Lila, Nunik, Risma, Ervin, Aty dan Irva. Terima kasih atas segala perhatian, dukungan, bantuan, kebersamaan, keceriaan serta kekeluargaannya selama ini. Bersyukur sekali punya saudara seperjuangan seperti kalian.
9. Teman-teman BPS angkatan 9 : Mbak Kiki, Mbak Ayu, Ika, Tiara, Mbak Dewi, Mas Agung, Mas Dinu, Mas Arif, Mas Bambang, Bang Node, Suko, Leman dan Bayu. Terima kasih atas kebersamaan dan kekompakannya selama Penulis menyelesaikan studi di ITS. Bersyukur dapat bertemu dengan kalian dan semoga dapat bertemu di lain kesempatan.
10. Mas Ahmad Syahrul Choir, atas bantuannya pada pengolahan data. Teman dan kerabat lain serta semua pihak yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa tesis ini jauh dari sempurna. Oleh karena itu, kritik maupun saran yang bersifat membangun diharapkan demi perbaikan tesis ini. Akhirnya, Penulis berharap semoga tesis ini bermanfaat untuk semua pihak yang memerlukan.

Surabaya, Januari 2017

Penulis



## DAFTAR ISI

ABSTRAK .....	v
ABSTRACT .....	vii
KATA PENGANTAR .....	ix
DAFTAR ISI .....	xi
DAFTAR TABEL .....	xv
DAFTAR GAMBAR .....	xvii
DAFTAR LAMPIRAN .....	xix
BAB 1     PENDAHULUAN .....	1
1.1     Latar Belakang .....	1
1.2     Rumusan Masalah .....	4
1.3     Tujuan Penelitian .....	6
1.4     Manfaat Penelitian .....	7
1.5     Batasan Masalah .....	7
BAB 2     TINJAUAN PUSTAKA .....	9
2.1     Analisis <i>Cluster</i> .....	9
2.2     Perkembangan Metode Analisis <i>Cluster</i> .....	10
2.3     Distribusi Probabilitas .....	13
2.3.1     Distribusi Multinomial .....	13
2.3.2     Distribusi Normal Multivariat .....	14
2.3.3     Distribusi <i>t</i> Multivariat .....	15
2.4 <i>Model Based-Clustering</i> .....	16
2.4.1     Model <i>Finite Mixture</i> Normal Multivariat .....	19
2.4.3     Model <i>Finite Mixture t</i> Multivariat .....	20

2.5	Penaksiran Parameter pada Model <i>Finite Mixture</i> dengan Metode <i>Maximum Likelihood</i> .....	21
2.6	Algoritma EM dan ECM .....	23
2.7	Pemilihan Model Terbaik .....	25
2.7.1	<i>Integrated Completed Likelihood</i> (ICL) .....	25
2.7.2	<i>Minimum Message Length</i> (MML) .....	26
2.8	Deteksi <i>Outlier</i> Multivariat .....	27
2.9	Skewness dan Kurtosis .....	29
2.10	Uji <i>Manova</i> Satu Arah .....	30
2.11	Survei Angkatan Kerja Nasional (Sakernas) .....	33
2.11.1	Kerangka Sampel .....	34
2.11.2	Desain Sampel .....	35
2.11.3	Konsep Umum Ketenagakerjaan .....	36
2.12	Indikator Pasar Tenaga Kerja ( <i>Key Indicators of The Labor Market</i> KILM) .....	38
2.12.1	KILM 1. Tingkat Partisipasi Angkatan Kerja (TPAK) .....	39
2.12.2	KILM 2. Rasio Penduduk yang Bekerja Terhadap Jumlah Penduduk ( <i>Employment to Population Ratio-EPR</i> ) .....	40
2.12.3	KILM 3. Penduduk yang Bekerja Menurut Status Pekerjaan Utama .....	41
2.12.4	KILM 4. Penduduk yang Bekerja Menurut Sektor (Lapangan Usaha) .....	42
2.12.5	KILM 5. Pekerja Paruh Waktu .....	43
2.12.6	KILM 6. Penduduk yang Bekerja Menurut Jam Kerja .....	43
2.12.7	KILM 7. Penduduk yang Bekerja di Sektor Informal .....	43
2.12.8	KILM 8. Pengangguran .....	43
2.12.9	KILM 11. Pengangguran dan Pendidikan .....	45

2.12.10	KILM 12. Setengah Penganggur .....	45
BAB 3	METODOLOGI PENELITIAN .....	47
3.1	Sumber Data .....	47
3.2	Variabel Penelitian .....	47
3.3	Tahapan Penelitian .....	50
BAB 4	HASIL DAN PEMBAHASAN .....	53
4.1	Subset Data Penelitian Menurut Indikator Pasar Tenaga Kerja Indonesia yang Memenuhi Asumsi Distribusi $t$ Multivariat .....	53
4.1.1	Deteksi <i>Outlier</i> Multivariat .....	53
4.1.2	Pengujian Asumsi Distribusi .....	55
4.2	Pengelompokan Provinsi di Indonesia Menurut Indikator Pasar Tenaga Kerja .....	57
4.2.1	Pengelompokan Provinsi di Indonesia menggunakan <i>Model- Based Clustering</i> dengan Kriteria <i>Integrated Completed Likelihood</i> (MBC-ICL) .....	57
4.2.2	Pengelompokan Provinsi di Indonesia menggunakan <i>Robust Model Based Clustering</i> dengan Kriteria <i>Minimum Message Length</i> (RMBC-MML) .....	60
4.3	Kelompok Optimal MBC-ICL dan RMBC-MML .....	64
4.4	Uji Kesamaan Kelompok .....	66
4.5	Analisis Pasar Tenaga Kerja Indonesia .....	67
4.5.1	Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Lapangan Usaha ( $k_2$ ) .....	68
4.5.2	Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan ( $k_3$ ) .....	72

4.5.3	Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur Menurut Pendidikan ( $k_4$ ) .....	76
4.5.4	Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator EPR, Pekerja Rentan, dan Pekerja Sektor Informal ( $k_5$ ) .....	80
BAB 5	KESIMPULAN DAN SARAN .....	83
5.1	Kesimpulan.....	83
5.2	Saran .....	84
DAFTAR PUSTAKA.....		87
BIOGRAFI PENULIS.....		145

## DAFTAR TABEL

Tabel 2.1	Struktur Data Kelompok ke- $g$ pada Model <i>Finite Mixture</i> .....	18
Tabel 2.2	<i>Multivariate Analysis of Variance</i> (Manova).....	30
Tabel 2.3	Pendekatan F untuk kriteria pengujian Manova.....	32
Tabel 3.1	Variabel Penelitian .....	47
Tabel 3.2	Kombinasi Variabel pada Subset data KILM.....	49
Tabel 4.1	Jumlah Provinsi Terdeteksi Sebagai Outlier Pada Kuantil 90 Persen .....	55
Tabel 4.2	Jumlah Kelompok Optimal MBC-ICL.....	59
Tabel 4.3	Anggota Kelompok Provinsi di Indonesia Berdasarkan Subset Data $k_2$ 0212 Menggunakan MBC-ICL .....	59
Tabel 4.4	Jumlah Kelompok Optimal RMBC-MML .....	62
Tabel 4.5	Anggota Kelompok Provinsi di Indonesia Berdasarkan Subset Data $k_2$ 0212 Menggunakan MBC-MML.....	63
Tabel 4.6	Klasifikasi Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Lapangan Usaha ( $k_2$ 0214) Menggunakan MBC-ICL dan MBC-MML (Kelompok Optimal) .....	64
Tabel 4.7	Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Lapangan Usaha ( $k_2$ 0214) Menggunakan MBC-ICL dan MBC-MML (Kelompok Optimal).....	65
Tabel 4.8	Hasil Uji Kesamaan Kelompok (Uji Manova) Subset Data $k_2$ 0815, $k_3$ 0815, $k_4$ 0815, dan $k_5$ 0815 dengan Statistik Uji <i>Pillai's Trace</i> .....	67
Tabel 4.9	Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Lapangan Usaha, Agustus 2015 .....	69
Tabel 4.10	Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Lapangan Usaha, Agustus 2015 .....	71

Tabel 4.11	Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Status, dan TPT Menurut Pendidikan, Agustus 2015 .....	73
Tabel 4.12	Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan, Agustus 2015.....	75
Tabel 4.13	Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur, Agustus 2015 .....	77
Tabel 4.14	Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur, Agustus 2015.....	79
Tabel 4.15	Pengelompokan Provinsi di Indonesia Berdasarkan Indikator EPR ( $X_2$ ), Pekerja Rentan ( $X_8$ ), dan Pekerja Sektor Informal ( $X_{18}$ ), Agustus 2015 .....	81
Tabel 4.16	Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator EPR, Pekerja Rentan, dan Pekerja Sektor Informal, Agustus 2015 .....	82

## DAFTAR GAMBAR

Gambar 2.1	Contoh Bentuk <i>Cluster</i> yang Merepresentasikan Struktur Matriks Varians Kovarians pada <i>Model-Based Clustering</i> .....17
Gambar 2.2	Bagan Konsep Dasar Tenaga Kerja .....36
Gambar 3.1	Diagram Alir Tahapan Penelitian .....51
Gambar 4.1	<i>Plot</i> Jarak Mahalanobis Terhadap Jarak <i>Robust</i> Subset Data $k_1$ 0212 .....54
Gambar 4.2	<i>Marginal Contour Plot</i> Subset Data $k_2$ 0212 .....58
Gambar 4.3	Persentase Bekerja Menurut Lapangan Usaha, Agustus 2014-2015 .....68
Gambar 4.4	<i>Plot</i> Keanggotaan <i>Cluster</i> Subset Data $k_2$ 0212- $k_2$ 0815 .....69
Gambar 4.5	Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Lapangan Usaha ( $k_2$ ), Agustus 2015 .....70
Gambar 4.6	<i>Plot</i> Keanggotaan <i>Cluster</i> Subset Data $k_3$ 0212- $k_3$ 0815 .....72
Gambar 4.7	Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan ( $k_3$ ), Agustus 2015 .....74
Gambar 4.8	<i>Plot</i> Keanggotaan <i>Cluster</i> Subset Data $k_4$ 0212- $k_4$ 0815 .....76
Gambar 4.9	Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur, Agustus 2015 .....78
Gambar 4.10	Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator EPR, Pekerja Rentan, dan Pekerja Sektor Informal, Agustus 2015 .....80



## DAFTAR LAMPIRAN

Lampiran 1.	Tata Nama <i>Model-Based Clustering</i> dalam <i>Package Teigen Software R</i> .....	93
Lampiran 2.	Tata Nama Kombinasi Variabel Subset Data KILM .....	94
Lampiran 3.	Variabel Penelitian .....	95
Lampiran 4.	Data Indikator Pasar Tenaga Kerja Sakernas Februari 2012 ( $X_1$ sampai $X_{25}$ ) .....	97
Lampiran 5	Data Indikator Pasar Tenaga Kerja Sakernas Februari 2013 ( $X_1$ sampai $X_{25}$ ) .....	99
Lampiran 6	Data Indikator Pasar Tenaga Kerja Sakernas Februari 2014 ( $X_1$ sampai $X_{25}$ ) .....	101
Lampiran 7	Data Indikator Pasar Tenaga Kerja Sakernas Februari 2015 ( $X_1$ sampai $X_{25}$ ) .....	103
Lampiran 8	Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2012 ( $X_1$ sampai $X_{25}$ ) .....	105
Lampiran 9	Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2013 ( $X_1$ sampai $X_{25}$ ) .....	107
Lampiran 10	Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2014 ( $X_1$ sampai $X_{25}$ ) .....	109
Lampiran 11	Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2015 ( $X_1$ sampai $X_{25}$ ) .....	111
Lampiran 12.	Output deteksi <i>outlier</i> subset data KILM $k_1$ 0212 menggunakan <i>software R</i> .....	113
Lampiran 13.	Provinsi <i>Outlier</i> Pada Subset Data $k_1$ Sakernas Februari 2012 Sampai Sakernas Agustus 2015 .....	114
Lampiran 14.	Provinsi <i>Outlier</i> Pada Subset Data $k_2$ Sakernas Februari 2012 Sampai Sakernas Agustus 2015 .....	115
Lampiran 15.	Provinsi <i>Outlier</i> Pada Subset Data $k_3$ Sakernas Februari 2012 Sampai Sakernas Agustus 2015 .....	116

Lampiran 16.	Provinsi <i>Outlier</i> Pada Subset Data $k_4$ Sakernas Februari 2012 Sampai Sakernas Agustus 2015.....	117
Lampiran 17.	Provinsi <i>Outlier</i> Pada Subset Data $k_5$ Sakernas Februari 2012 Sampai Sakernas Agustus 2015.....	118
Lampiran 18.	<i>Output Mardia test</i> Untuk Uji Normal Multivariat Subset Data KILM $k_1$ 0212 – $k_1$ 0815.....	119
Lampiran 19.	Hasil Uji Normal Multivariat Subset Data KILM $k_1$ 0215 – $k_5$ 0815.....	120
Lampiran 20.	<i>Script</i> dan Output Kelompok Optimal Subset Data KILM $k_2$ 0212 – $k_2$ 0815 dengan Kriteria ICL Menggunakan <i>Software R</i> .....	121
Lampiran 21.	Output Nilai ICL Subset Data KILM $k_2$ 0212.....	122
Lampiran 22.	Estimasi Parameter Subset Data KILM $k_2$ 0212 Pada Jumlah Kelompok ( $G$ ) = 2 , Model UUUC Menggunakan MBC- ICL .....	123
Lampiran 23.	Pengelompokan Provinsi di Indonesia Menurut Kombinasi Variabel $k_2$ (Persentase Bekerja Menurut Lapangan Usaha) Pada Jumlah Kelompok ( $G$ ) = 2 Menggunakan MBC- ICL.....	124
Lampiran 24.	Output Nilai ICL Subset Data KILM $k_5$ 0815.....	125
Lampiran 25.	Pengelompokan Provinsi di Indonesia Menurut Kombinasi Variabel $k_3$ (Bekerja Menurut Status, dan TPT Menurut Pendidikan) Pada Jumlah Kelompok Optimal Menggunakan MBC- ICL .....	126
Lampiran 26.	Output Nilai ICL Subset Data KILM $k_3$ 0815.....	127
Lampiran 27.	Pengelompokan Provinsi di Indonesia Menurut Kombinasi Variabel $k_4$ (Persentase Bekerja Menurut Jam Kerja, dan Persentase Pekerja Setengah Penganggur) Pada Jumlah Kelompok ( $G$ ) = 2 Menggunakan MBC- ICL .....	128
Lampiran 28.	Output Nilai ICL Subset Data KILM $k_4$ 0815.....	129
Lampiran 29.	<i>Script Model-Based Clustering</i> Dengan Kriteria ICL Pada <i>Software R</i> .....	130
Lampiran 30	Program RMBC-MML.....	131

Lampiran 31.	<i>Main Script</i> RMBC-MML Kombinasi Variabel $k_1$ , $k_2$ , $k_3$ , $k_4$ dan $k_5$ Pada <i>Software</i> MATLAB, Sakernas Februari 2012 Sampai Dengan Februari 2015.....	139
Lampiran 32.	Output Kelompok Optimal RMBC-MML Pada Subset Data $k_2$ 0212, $k_2$ 0213, $k_2$ 0214, dan $k_5$ 0213 .....	141
Lampiran 33.	Perbandingan Keanggotaan Kelompok Optimal MBC-ICL Dengan RMBC-MML Pada Subset Data $k_2$ 0212, $k_2$ 0213, $k_2$ 0214, dan $k_5$ 0213 .....	142
Lampiran 34.	Hasil Pengelompokan dan Parameter Subset Data $k_5$ 0815 Menggunakan RMBC-MML .....	143

*Halaman ini sengaja dikosongkan*

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Penciptaan lapangan kerja menjadi salah satu fokus utama pembangunan. Tidak hanya sebatas “lapangan kerja”, tetapi bagaimana menciptakan “lapangan kerja layak” seperti yang dibahas pada Konferensi *Sustainable Development Goals* (SDGs) tanggal 16-17 Februari 2016 dengan tema “Agenda untuk Tujuan Pembangunan Berkelanjutan Menuju Kerja Layak untuk Semua”. Konferensi ini mengidentifikasi berbagai tantangan pasar tenaga kerja yang dihadapi Indonesia saat ini dan kebutuhan mendesak untuk mengatasi tantangan-tantangan tersebut secara komprehensif dan holistik. Upaya untuk menciptakan kesempatan kerja layak harus terus dilakukan sebagaimana tertuang dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN). Kesempatan kerja layak bukan saja menjadi perhatian pemerintah tetapi juga organisasi dunia seperti Organisasi Perburuhan Internasional (*International Labour Organization/ILO*), melalui Program Kerja Layak Nasional 2012-2015 (*Indonesia Decent work Country Programme/DWCP*).

Survei Angkatan Kerja Nasional (Sakernas) Februari 2016 mencatat Tingkat Pengangguran Terbuka (TPT) Indonesia mencapai 5,50 persen. Ini berarti dari 100 orang angkatan kerja di Indonesia, terdapat 5 atau 6 orang yang masuk kategori penganggur. Selain itu, *share* penganggur umur muda terhadap total penganggur juga masih cukup tinggi yaitu 53,12 persen. Dari 100 orang penganggur terdapat 53 orang penganggur yang berumur antara 15 sampai 24 tahun (BPS, 2016). Keadaan ini semakin mempertegas diperlukannya perhatian dan upaya yang lebih besar lagi untuk mendukung kesempatan kerja di Indonesia terutama bagi pekerja umur muda. BPS sudah menyajikan informasi yang cukup lengkap mengenai kondisi pasar tenaga kerja di Indonesia. Dalam satu tahun sedikitnya terdapat 4 publikasi yang menggambarkan secara rinci mengenai ketenagakerjaan Indonesia, seperti publikasi : “Statistik Mobilitas Penduduk dan

Tenaga Kerja”, “Keadaan Pekerja di Indonesia”, “Keadaan Angkatan Kerja” dan “Indikator Pasar Tenaga Kerja”. Akan tetapi, publikasi-publikasi tersebut umumnya menyajikan informasi melalui tabel-tabel, grafik dan analisis yang bersifat deskriptif. Belum pernah disajikan informasi berdasarkan analisis *clustering* (pengelompokan) provinsi-provinsi di Indonesia, sehingga bisa diketahui provinsi mana saja yang memiliki kemiripan karakteristik menurut beberapa indikator pasar tenaga kerja.

ILO sudah menetapkan bahwa untuk mengukur pasar tenaga kerja suatu negara dibutuhkan banyak indikator (multivariat). Perkembangan pasar tenaga kerja tidak cukup hanya dilihat dari satu dimensi atau satu indikator (univariat) saja. Sebagai contoh, Tingkat Partisipasi Angkatan Kerja (TPAK) Papua selalu menduduki angka tertinggi selama 2012-2015 yaitu di atas 78 persen. Begitu juga dengan Bali, menempati posisi kedua dengan TPAK di atas 75 persen. Jika dilihat dari indikator TPAK, kedua provinsi ini memiliki karakteristik yang sama dan akan berada di kelompok yang sama pula, yaitu TPAK tinggi. Suatu angka yang mengindikasikan bahwa kedua provinsi ini memiliki stok tenaga kerja yang cukup besar. Namun, jika dikaji dari sisi indikator Tingkat Pengangguran Terbuka (TPT) menurut pendidikan, Papua dan Bali memiliki karakteristik yang berbeda. Kondisi Februari 2015, Provinsi Papua berada pada posisi dengan TPT pendidikan tinggi sebesar 8,44 persen, jauh di atas angka nasional 5,86 persen. Sedangkan TPT pendidikan tinggi Provinsi Bali hanya 1,43 persen. Oleh karena itu, dibutuhkan metode *clustering* yang tepat untuk mengakomodir persamaan dan perbedaan karakteristik tenaga kerja provinsi-provinsi di Indonesia jika dilihat dari banyak variabel (multivariat).

*Model-based clustering* merupakan salah satu teknik *clustering* yang cukup berkembang dan dianggap memiliki kelebihan dibanding metode *clustering* klasik lainnya. Istilah *model-based clustering* (MBC) pertama kali digunakan oleh Banfield & Raftery (1993) untuk menggambarkan sebuah pendekatan *clustering* dimana suatu kelompok pada populasi diidentifikasi berdasar distribusi probabilitas dan keseluruhan populasi dimodelkan sebagai sebuah *finite mixture* distribusi. Sehingga dapat dikatakan metode MBC adalah teknik *clustering* berdasarkan model probabilitas. Menurut Bouveyron &

Brunet Saumard (2014), metode *clustering* dalam kerangka model probabilitas memberikan struktur yang tetap tentang kelompok-kelompok melalui distribusi probabilitas kelompok tersebut. Metode ini dianggap lebih baik daripada metode *cluster* yang umum digunakan selama ini, yang mengukur kesamaan antar objek melalui ukuran jarak, ukuran korelasi dan ukuran asosiasi. Pengukuran kemiripan antar objek dengan menggunakan jarak akan sangat sulit dilakukan jika kondisi objek yang ada saling tumpang tindih.

Penggunaan model *finite mixture* dalam metode *cluster* telah ada sejak tahun 1960-an. Dimulai dengan karya Wolfe (1965), Edwards & Cavalli-Sforza (1965), Day (1969), McLachlan (1982), Titterington, Smith & Makov (1985), Banfield & Raftery (1993). Dasgupta & Raftery (1998), McLachlan & Peel (2000), Fraley & Raftery (2002), Cuesta & Albertos (2008) dan McLachlan & Wang (2012). Sebagian besar penelitian tersebut didasarkan pada model *mixture normal*. Penelitian terbaru yang menggunakan teknik MBC dilakukan oleh Damayanti (2015) yang mengelompokkan provinsi-provinsi di Indonesia menurut Capaian Pembangunan Berkelanjutan dan Siagian (2014) yang mengelompokkan Kabupaten/ Kota di Indonesia menurut Tingkat Kerawanan Sosial. Mengacu penelitian McLachlan & Peel pada tahun 2000, penelitian Damayanti (2015) dan Siagian (2014) didasarkan pada MBC dengan distribusi *t* multivariat. Menurut McLachlan & Peel (2000) untuk kasus data mengandung *outlier*, penggunaan asumsi *mixture* berdistribusi *t* multivariat akan memberikan hasil *clustering* yang lebih *robust* dalam mengatasi efek *outlier* pada estimasi parameter.

Penelitian ini akan menerapkan metode penelitian yang dilakukan oleh Siagian (2014) dan Damayanti (2015), untuk mengelompokkan provinsi di Indonesia menurut indikator pasar tenaga kerja. Potensi tenaga kerja dan kondisi geografis yang berbeda-beda antar provinsi terkadang memuat data ekstrim (*outlier*). Keberadaan *outlier* bisa menjadi salah satu penyebab analisis yang dihasilkan tidak menggambarkan dengan baik keadaan di lapangan. Damayanti (2015) menggunakan kriteria *Integrated Completed Likelihood* (ICL) untuk mengelompokkan 33 provinsi di Indonesia menurut 10 variabel indikator pembangunan tahun 2011, sedangkan Siagian (2014) menggunakan kriteria *Minimum Message Length* (MML) yang terbukti lebih baik dibandingkan kriteria



*Bayesian Information Criterion* (BIC) untuk mengelompokkan 497 kabupaten/kota berdasarkan 10 variabel data kerawanan sosial tahun 2010. Keduanya menggunakan data riil bersumber dari Badan Pusat Statistik (BPS) dan diasumsikan berdistribusi  $t$  multivariat dan memuat *outlier*. Kedua kriteria ICL dan MML akan digunakan pada penelitian ini untuk mengelompokkan 33 provinsi di Indonesia menurut data indikator pasar tenaga kerja yang diduga mengandung *outlier* dan tidak berdistribusi normal. Jika penelitian Damayanti dan Siagian melakukan pengelompokan pada satu tahun (periode) pendataan saja, maka tesis ini akan menerapkan kedua kriteria tersebut pada 8 periode pendataan. Pengelompokan (*clustering*) melalui metode “*Model-Based Clustering* dengan Distribusi  $t$  Multivariat Menggunakan Kriteria *Integrated Completed Likelihood* dan *Minimum Message Length*” diharapkan bisa menghasilkan pengelompokan yang mampu memberikan informasi yang tepat mengenai pasar tenaga kerja provinsi di Indonesia dari tahun 2012 sampai dengan 2015.

## 1.2 Rumusan Masalah

Pada prinsipnya pengembangan metode statistik diperlukan karena dunia kerja menuntut pengumpulan data yang lengkap dan analisis informasi yang luas mengenai pasar tenaga kerja Indonesia. BPS sejak tahun 2012 secara rutin menyampaikan hasil pencacahan Sakernas melalui publikasi “Indikator Pasar Tenaga Kerja” berdasarkan “*Key Indicators of Labor Market KILM*” edisi ke lima yang ditetapkan oleh ILO. Dari 20 indikator KILM oleh ILO, BPS bisa menyajikan 13 indikator KILM yang dijelaskan oleh lebih dari 30 variabel indikator pasar tenaga kerja. Jika pada penelitian Siagian (2014) dan Damayanti (2015) dilakukan *clustering* objek pada satu periode data penelitian, maka pada penelitian ini akan dilakukan *clustering* objek yaitu provinsi-provinsi di Indonesia dengan menggunakan 8 periode data penelitian Sakernas bulan Februari dan Agustus tahun 2012 sampai dengan 2015. Penggunaan variabel yang sama pada periode yang berbeda memungkinkan hasil pengelompokan yang berbeda. Apalagi jika variabel yang digunakan berbeda pada periode yang berbeda, maka sangat mungkin hasil pengelompokan akan berbeda. Hasil pengelompokan yang berbeda-beda atau tidak stabil akan menyulitkan pengambilan keputusan. Melalui

eksplorasi variabel-variabel indikator pasar tenaga kerja diharapkan akan diperoleh variabel yang bisa menghasilkan pengelompokan yang stabil, yaitu hasil pengelompokan yang hampir konsisten dari waktu ke waktu dengan asumsi tidak ada kejadian luar biasa yang dapat merubah komposisi pengelompokan secara drastis.

*Model-based clustering* (MBC) umumnya dilakukan untuk mengelompokkan objek dengan satu periode data penelitian (*cross-section*). MBC untuk mengelompokkan objek dengan beberapa *series* data penelitian telah mengalami banyak perkembangan. Juarez dan Steel (2010) melakukan pengelompokan dengan metode MBC untuk data panel, sedangkan Fruhwirth-Schnatter (2011) melakukan pengelompokan dengan metode MBC untuk data *timeseries*. Meskipun penelitian ini memuat 8 *series* data indikator pasar tenaga kerja, akan tetapi pengelompokan provinsi di Indonesia akan dilakukan secara *cross-section*. Penelitian ini tidak ditujukan untuk mendapatkan satu pengelompokan provinsi dari 8 *series* data pencacahan Sakernas seperti *clustering* pada data panel ataupun data *timeseries*. Akan tetapi, penelitian ini akan mendapatkan pengelompokan provinsi untuk masing-masing periode pencacahan Sakernas yaitu februari 2012, agustus 2012, februari 2013, agustus 2013, februari 2014, agustus 2014, februari 2015 dan agustus 2015. Dengan demikian diharapkan bisa diketahui apakah terdapat perbedaan kelompok yang terbentuk untuk setiap periode pencacahan Sakernas.

Distribusi *finite mixture* yang banyak digunakan dalam penelitian *model-based clustering* adalah distribusi normal multivariat karena relatif sederhana dalam perhitungannya (Banfield & Raftery, 1993). Akan tetapi banyak distribusi data multivariat yang tidak mengikuti distribusi normal akibat keberadaan *outlier*. Terdapat dampak tidak baik jika memaksakan asumsi model *mixture* normal multivariat pada data mengandung nilai ekstrim karena akan terjadi *over-estimate* jumlah kelompok (Cozzini, Asra & Montana, 2013). *Over-estimate* terjadi karena kelompok tambahan dibutuhkan untuk mencakup distribusi *heavy tail* yang menggambarkan karakteristik kelompok tertentu (Melnikov & Maitra, 2010). Penggunaan metode MBC pada penelitian ini dikhususkan untuk data indikator pasar tenaga kerja yang memenuhi asumsi distribusi *t* multivariat. Pengelompokan

melalui metode MBC dengan distribusi  $t$  multivariat diharapkan bisa menghasilkan penduga yang *robust* untuk mengelompokkan provinsi berdasarkan data ketenagakerjaan yang cenderung bervariasi dan memuat *outlier*.

Penaksiran parameter dan pemilihan model terbaik merupakan dua proses utama pada metode *clustering* model *finite mixture*. Untuk mendapatkan *Robust Model-Based Clustering* (RMBC) terbaik dan jumlah kelompok yang optimal, berbagai kriteria telah dikembangkan. Peel & McLachlan (2000) menggunakan metode *Maximum Likelihood* (ML) untuk estimasi parameter serta *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) untuk seleksi model. Damayanti (2015) juga menggunakan metode ML untuk estimasi parameter, sedangkan untuk seleksi model terbaik menggunakan kriteria *Integrated Completed Likelihood* (ICL). Siagian (2014) mengembangkan metode *Maximum Penalized Likelihood* (MPL) untuk estimasi parameter dan penalti *Minimum Message Length* (MML) untuk memilih model terbaik. Seperti telah disampaikan di latar belakang, tesis ini akan menerapkan kedua metode yang dilakukan oleh Damayanti (2015) dan Siagian (2014) untuk mengelompokkan provinsi di Indonesia berdasarkan indikator pasar tenaga kerja.

Berikut ini beberapa permasalahan yang akan dibahas dalam penelitian, menggunakan indikator pasar tenaga kerja Februari dan Agustus tahun 2012 sampai dengan 2015.

- 1) Bagaimana memilih subset data penelitian menurut indikator pasar tenaga kerja Indonesia yang memenuhi asumsi distribusi  $t$  multivariat?
- 2) Bagaimana membentuk kelompok wilayah (provinsi) menurut indikator pasar tenaga kerja menggunakan *model-based clustering*  $t$  multivariat dan kriteria pemilihan model terbaik menggunakan kriteria ICL dan MML?
- 3) Berdasarkan kelompok yang terbentuk, bagaimana analisis pasar tenaga kerja Indonesia ?

### 1.3 Tujuan Penelitian

Berdasarkan perumusan masalah di atas, tujuan utama penelitian ini adalah mengembangkan metode statistik *model-based clustering* dengan

membandingkan kriteria ICL dan MML, untuk memilih model terbaik pada model *finite mixture t multivariate*. Secara khusus tujuan penelitian adalah:

- 1) Memilih subset data penelitian menurut indikator pasar tenaga kerja Indonesia yang memenuhi asumsi distribusi *t* multivariat.
- 2) Membentuk kelompok provinsi di Indonesia menurut indikator pasar tenaga kerja menggunakan *model-based clustering t* multivariat dan kriteria pemilihan model terbaik menggunakan kriteria ICL dan MML
- 3) Menganalisis pasar tenaga kerja Indonesia berdasarkan kelompok yang terbentuk.

#### **1.4 Manfaat Penelitian**

Hasil penelitian ini dapat diterapkan untuk pengelompokan (*clustering*) objek berdasarkan data riil yang umumnya banyak memuat data *outlier*. Juga diharapkan bisa menjadi informasi tambahan bagi BPS selaku instansi pemerintah yang secara rutin melakukan Survei Angkatan Kerja Nasional (Sakernas). Penelitian ini menggunakan series data Sakernas dari tahun 2012 sampai 2015, sehingga hasil analisis penelitian ini bisa mendukung Indonesia *Decent work Country Programme* 2012-2015 oleh ILO, selaku organisasi internasional yang sudah sejak lama peduli dengan perkembangan pasar tenaga kerja Indonesia. Pada akhirnya, semoga hasil penelitian ini bisa memberikan masukan penting bagi semua pengguna data ketenagakerjaan dan bagi pemerintah dalam pengambilan kebijakan yang tepat, mendukung “Agenda Tujuan Pembangunan Berkelanjutan Menuju Kerja Layak untuk Semua”.

#### **1.5 Batasan Masalah**

Berdasarkan latar belakang dan rumusan masalah, perlu dilakukan batasan cakupan dan batasan masalah penelitian, sehingga tujuan yang diharapkan bisa tercapai. Beberapa batasan masalah yang digunakan pada penelitian ini, seperti sebagai berikut :

1. Pengelompokan dilakukan menurut 33 Provinsi di Indonesia, tidak termasuk Kalimantan Utara. Sebagai provinsi termuda, series data Provinsi Kalimantan

Utara belum lengkap. Data KILM Kalimantan Utara tahun 2012-2014 tidak tersedia.

2. Dari 20 indikator KILM yang ditetapkan ILO, penelitian ini menggunakan 10 indikator KILM yang terdiri dari 25 indikator/variabel pasar tenaga kerja berdasarkan hasil pencacahan Sakernas bulan Februari dan Agustus tahun 2012 sampai dengan 2015 (seperti pada Tabel 3.1). Dengan asumsi bahwa data Indikator Pasar Tenaga Kerja (KILM) Indonesia cenderung memuat *outlier*, maka pengelompokan provinsi di Indonesia dilakukan berdasarkan indikator-indikator KILM yang memenuhi asumsi distribusi  $t$  multivariat.
3. Pengelompokan provinsi di Indonesia dilakukan pada setiap periode pencacahan Sakernas bulan Februari dan Agustus tahun 2012 sampai dengan 2015. Ini berarti data dikelompokkan sebagai data *cross-section* untuk tiap periode Sakernas, bukan sebagai data panel 2012-2015 atau pun data *time-series* 2012-2015.

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1 Analisis Cluster**

Analisis *cluster* merupakan teknik mencari pola dalam kumpulan data dengan menggabungkan observasi ke dalam kelompok atau *cluster*. Tujuannya adalah menemukan sebuah pengelompokan optimal dimana pengamatan atau objek dalam setiap *cluster* serupa, tetapi antar *cluster* berbeda satu sama lain. Rencher (2002) menyebutkan bahwa analisis cluster secara fundamental berbeda dengan analisis klasifikasi. Dalam analisis klasifikasi, observasi dikelompokkan ke sejumlah kelompok atau populasi yang telah ditetapkan. Dalam analisis cluster, baik jumlah kelompok maupun kelompok itu sendiri belum pernah ditentukan atau diketahui sebelumnya. Analisis *cluster* berusaha menemukan pengelompokan secara alamiah dan hasil pengelompokan tersebut masuk akal atau sesuai dengan *sense* peneliti (Rencher, 2002). Bezdek dkk (1984) mendefinisikan analisis *cluster* sebagai sekumpulan metode untuk membagi suatu set data ke dalam subset-subset yang lebih kecil, yang saling terpisah satu sama lain.

Penggunaan analisis *cluster* dalam analisis data telah berkembang sangat pesat untuk berbagai keperluan. Mulai dari masalah segmentasi gambar, pengelompokan dokumen, akses informasi berdasarkan topik, segmentasi pelanggan untuk pemasaran, hingga penelitian data genetik. Menurut Jain (2010), analisis *cluster* dipergunakan untuk tiga tujuan utama:

- i. Metode untuk penentuan struktur dalam rangka mendapatkan pola data, membangkitkan hipotesis, mendeteksi adanya penyimpangan, dan mengidentifikasi ciri tertentu yang menonjol.
- ii. Metode klasifikasi alamiah untuk menentukan derajat kesamaan antara bentuk-bentuk makhluk hidup (hubungan filogenik).
- iii. Metode untuk mendapatkan struktur data yang lebih ringkas dan terorganisasi.

Seiring dengan perkembangan ilmu pengetahuan, kebutuhan terhadap data dengan dimensi yang semakin besar menjadi tidak terelakkan. Untuk mengatasi kerumitan

yang ditimbulkan dalam pengelolaan data dengan jumlah yang besar, diperlukan adanya proses penyederhanaan melalui pengorganisasian dan pengkategorian data. Manusia cenderung untuk meringkas kumpulan data dengan dimensi yang besar (mutivariat) ke dalam kelompok-kelompok yang lebih kecil. Pada beberapa kasus, data dengan karakteristik yang sama memiliki kecenderungan alamiah untuk membentuk kelompok tersendiri. Namun pada saat dimensi data cukup besar, maka upaya untuk mendapatkan kelompok data dengan karakteristik yang sama menjadi semakin sulit. Pertumbuhan volume serta variasi data yang semakin besar memerlukan metode pengelompokan yang dapat mempartisi set data dengan ketepatan yang tinggi. Dalam analisis multivariat, salah satu metode yang banyak digunakan untuk mengatasi permasalahan tersebut adalah analisis *cluster*. Perkembangan analisis *cluster* mendapatkan sumbangan dari berbagai disiplin ilmu seperti taksonomi, ilmu sosial, psikologi, biologi, statistika, matematika, teknik, komputasi, kesehatan, dan ilmu lainnya yang melibatkan adanya kegiatan pengumpulan data. Penggunaan metode pengelompokan data pertama kali muncul tahun 1954 pada artikel sebuah jurnal yang terkait dengan data antropologi (Jain, 2010).

## **2.2 Perkembangan Metode Analisis *Cluster***

Proses pengelompokan dipandang sebagai suatu pendekatan yang baik untuk menemukan kesamaan pada data, dan menglompokkan data yang memiliki kesamaan ke dalam suatu kelompok tertentu (Hammouda dan Karray, 2000). Ukuran kesamaan adalah hal yang fundamental dalam analisis *cluster*. Kesamaan antar objek merupakan ukuran korespondensi antar objek. Ada tiga ukuran yang dapat digunakan, yaitu ukuran korelasi, ukuran jarak, dan ukuran asosiasi.

### **a) Ukuran Korelasi**

Ukuran ini dapat diterapkan pada data dengan skala metrik, tetapi jarang digunakan karena titik beratnya pada nilai suatu pola tertentu. Padahal titik berat analisis *cluster* adalah besarnya objek. Kesamaan antar objek dapat dilihat dari koefisien korelasi antar pasangan objek yang diukur dengan beberapa variabel.



b) Ukuran Jarak

Merupakan ukuran kesamaan yang paling sering digunakan. Diterapkan untuk data berskala metrik. Sesungguhnya ukuran jarak merupakan ukuran ketidakmiripan, dimana jarak yang besar menunjukkan sedikit kesamaan, sedangkan jarak yang pendek/kecil menunjukkan bahwa suatu objek makin mirip dengan objek lain. Fokus ukuran jarak terletak pada besarnya nilai. *Cluster* berdasarkan ukuran korelasi bisa saja tidak memiliki kesamaan nilai, tetapi memiliki kesamaan pola. Sedangkan *cluster* berdasarkan ukuran jarak lebih memiliki kesamaan nilai meskipun polanya berbeda. Beberapa tipe ukuran jarak antara lain jarak *euclidean*, jarak *city-block (manhattan)*, dan jarak *mahalanobis*.

c) Ukuran Asosiasi

Ukuran asosiasi dipakai untuk mengukur data berskala nonmetrik (nominal atau ordinal).

Rencher (2002) membagi pengelompokan objek melalui dua pendekatan umum yaitu metode hirarki dan metode non hirarki.

1. Metode Hirarki (*Hierarchical Methods*)

Tipe dasar dalam metode ini adalah aglomerasi dan pemecahan. Dalam metode aglomerasi tiap observasi pada mulanya dianggap sebagai *cluster* tersendiri, sehingga terdapat *cluster* sebanyak jumlah observasi. Kemudian dua *cluster* yang terdekat kesamaannya digabung menjadi suatu *cluster* baru, sehingga jumlah *cluster* berkurang satu pada tiap tahap. Sebaliknya pada metode pemecahan dimulai dari satu *cluster* besar yang mengandung seluruh observasi. Selanjutnya observasi-observasi yang paling berbeda dipisah dan dibentuk *cluster-cluster* yang lebih kecil. Proses ini dilakukan hingga tiap observasi menjadi *cluster* sendiri-sendiri. Hal penting dalam metode hirarki adalah bahwa hasil pada tahap sebelumnya selalu bersarang di dalam hasil pada tahap berikutnya, membentuk sebuah pohon. Ada lima metode aglomerasi dalam pembentukan *cluster* hirarki, yaitu pautan tunggal (*single linkage*), pautan lengkap (*complete linkage*), pautan rata-rata (*average linkage*), metode *ward* (*ward's method*), dan metode *centroid* (*centroid method*).

## 2. Metode Nonhirarki (*Nonhierarchical Methods*)

Perhatian utama dalam metode nonhirarki adalah bagaimana memilih bakal atau inisial *cluster*. Pemilihan inisial *cluster* berpengaruh terhadap hasil akhir analisis *cluster*. Inisial *cluster* pertama adalah observasi pertama dalam set data tanpa *missing value*. Inisial kedua adalah observasi lengkap berikutnya yang dipisahkan dari inisial pertama oleh jarak minimum khusus. Berikut tiga teknik pengelompokan pada metode nonhirarki.

### 1. *K-Means*

Algoritma *Lloyd* atau dikenal sebagai *K-means* memiliki sejarah panjang selama lebih dari 50 tahun sejak penggunaannya pertama kali oleh Steinhaus pada 1956 (Jain, 2010). Meskipun bukan merupakan metode baru, algoritma partisional ini masih merupakan salah satu yang paling banyak digunakan untuk keperluan pengelompokan terkait dengan kemudahan dan efisiensinya dalam penerapan, serta keberhasilannya dalam penerapan secara empiris untuk berbagai kasus analisis *cluster*. Salah satu masalah utama dalam optimasi *K-means* adalah mendapatkan suatu inialisasi yang menghasilkan solusi pengelompokan yang konvergen pada optimum global. Selama ini, proses untuk menghasilkan optimum global dilakukan dengan melakukan sejumlah pengulangan pada *K-means*, sehingga waktu komputasi yang diperlukan menjadi semakin panjang. Namun demikian, tetap tidak terdapat jaminan bahwa pengulangan *K-means* dengan inialisasi secara random akan menghasilkan optimum global, kecuali dilakukan dengan frekuensi yang sangat banyak. Untuk menghindari keadaan tersebut, diperlukan adanya suatu metode inialisasi cluster secara deterministik, yang tidak hanya mampu mendapatkan solusi optimum global, namun juga hasilnya bersifat unik.

### 2. *Density Estimation*

Pada metode ini, pengelompokan dilakukan dengan mencari wilayah yang memiliki kepadatan/ *density* yang tinggi. Estimasi *density* dilakukan menggunakan fungsi kernel dan tidak memerlukan asumsi.

### 3. *Mixtures of Distribution*

Pada metode ini, suatu kelompok pada populasi diidentifikasi berdasar distribusi probabilitas dan keseluruhan populasi dimodelkan sebagai sebuah *mixture* dari distribusi. Sehingga dapat dikatakan pengelompokan pada metode ini adalah pengelompokan berdasarkan model dari distribusi probabilitas.

## 2.3 Distribusi Probabilitas

Fenomena-fenomena yang dianggap acak atau tidak pasti sering dimodelkan sebagai hasil dari beberapa percobaan. Satu percobaan merupakan elemen ruang sampel, dan himpunan bagian dari ruang sampel tersebut dinamakan peristiwa atau kejadian (*events*). Kejadian-kejadian tersebut ditandai sebagai suatu peluang (probabilitas), yang bernilai antara 0 dan 1 yang menunjukkan seberapa sering *events* tersebut terjadi (Dekking dkk., 2005).

Distribusi peluang (distribusi probabilitas) sangat berguna untuk menganalisis terjadinya suatu peristiwa atau kejadian. Kejadian yang bersifat berhingga, objek sebarannya akan berbeda dengan kejadian yang tak berhingga. Objek dari sebaran peluang adalah variabel acak. Casella dan Berger (2002) mendefinisikan variabel acak sebagai fungsi yang memetakan kejadian-kejadian pada suatu ruang sampel ke dalam himpunan bilangan real. Sehubungan dengan pengujian objek percobaan, pemilihan sebaran akan mempermudah penghitungan peluang. Ditinjau dari objek kajian, distribusi probabilitas dibagi menjadi distribusi peluang diskrit dan distribusi peluang kontinu, dimana masing-masing memiliki beberapa jenis distribusi. Subbab ini membahas beberapa distribusi probabilitas yang digunakan pada penelitian ini.

### 2.3.1 Distribusi Multinomial

Sebuah variabel acak  $X$  dikatakan memiliki distribusi peluang diskrit jika ruang sampel  $X$  merupakan nilai yang bisa dihitung (*countable*). Atau variabel acaknya memiliki hasil percobaan berupa nilai *integer*. Distribusi multinomial adalah salah satu bentuk distribusi peluang diskrit. Distribusi multinomial merupakan perluasan dari distribusi binomial. Jika sebuah variabel acak binomial

berasal dari percobaan dengan 2 kemungkinan hasil maka variabel acak multinomial berasal dari percobaan dengan  $G$  kemungkinan hasil yaitu kejadian  $A_g$  dengan peluang  $\pi_g$  dimana  $g = 1, 2, \dots, G$  (Hogg & Tanis, 1997). Pada penelitian ini,  $G$  adalah jumlah kelompok yang terbentuk. Karena  $P(A_g) = \pi_g$  maka  $\sum_{g=1}^G \pi_g = 1$ . Percobaan dilakukan sebanyak  $n$  kali dan variabel acak  $X_g$  adalah banyaknya percobaan terjadi di  $A_g, g = 1, 2, \dots, G$ . Sehingga  $\sum_{g=1}^G X_g = n$  dan probabilitas  $X_1, X_2, \dots, X_{G-1}$  adalah

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_{G-1} = x_{G-1}) &= \frac{n!}{x_1! x_2! \dots x_{G-1}!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_{G-1}^{x_{G-1}} \\ &= n! \prod_{g=1}^{G-1} \frac{\pi_g^{x_g}}{x_g!} \end{aligned} \quad (2.1)$$

Variabel acak  $X_1, X_2, \dots, X_{G-1}$  dikatakan berdistribusi multinomial dengan parameter  $n, \pi_1, \pi_2, \dots, \pi_{G-1}$  (Forbes, C., dkk (2011)).

Rata-rata  $\mu_g = E[X_g] = n\pi_g$  dan varians  $\sigma_g^2 = Var[X_g] = n\pi_g(1 - \pi_g)$ .

Kovarians antara  $X_{g^*}$  dan  $X_g, g^* \neq g$  adalah  $\sigma_{g^*g} = Cov[X_{g^*}, X_g] = -n\pi_{g^*}\pi_g$ .

### 2.3.2 Distribusi Normal Multivariat

Suatu variabel acak  $X$  yang berdistribusi normal univariat dengan rata-rata  $\mu$  dan varians  $\sigma^2$  dinotasikan  $X \sim N(\mu, \sigma^2)$  memiliki bentuk fungsi kepadatan probabilitas :

$$f(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

dimana  $-\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$ . Fungsi kepadatan probabilitas normal multivariat adalah perluasan dari fungsi kepadatan probabilitas normal univariat untuk dimensi  $p \geq 2$ . Maka vektor variabel acak  $\mathbf{x} = [x_1 x_2 \dots x_p]^T$  berdistribusi normal multivariat dengan parameter  $\boldsymbol{\mu}$  dan  $\boldsymbol{\Sigma}$ , mempunyai fungsi kepadatan peluang :

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \quad (2.2)$$

dan dinotasikan dengan  $X \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  dimana :

$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$  adalah vektor rata-rata dan  $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ & \sigma_{22} & \dots & \sigma_{2p} \\ & & \ddots & \vdots \\ & & & \sigma_{pp} \end{pmatrix}$  adalah

matriks varians kovarians. Distribusi *finite mixture* yang banyak digunakan dalam penelitian *model-based clustering* adalah distribusi normal multivariat, karena relatif sederhana dalam perhitungannya (Banfield & Raftery, 1993).

### 2.3.3 Distribusi $t$ Multivariat

Distribusi  $t$  multivariat merupakan perluasan dari distribusi  $t$  univariat. Distribusi  $t$  multivariat mulai banyak digunakan dalam *model-based clustering*, karena banyak distribusi data multivariat yang tidak mengikuti distribusi normal akibat keberadaan *outlier*, sehingga distribusi data menjadi lebih landai. Kemudian dikembangkan distribusi  $t$  untuk data multivariat yang diakui memiliki kemampuan mengatasi *outlier* lebih mudah dibandingkan distribusi normal multivariat (Andrews., dkk, 2011 ).

Vektor variabel acak  $\mathbf{x} = [x_1 x_2 \dots x_p]^T$  berdistribusi  $t$  multivariat dengan derajat bebas  $v$ , vektor rata-rata  $\boldsymbol{\mu} = [\mu_1 \mu_2 \dots \mu_p]^T$  dan matriks varians kovarians  $\boldsymbol{\Sigma}$ , memiliki fungsi kepadatan peluang sebagai berikut (Kotz & Nadarajah, 2004) :

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{v+p}{2}\right)}{(\pi v)^{\frac{p}{2}} \Gamma\left(\frac{v}{2}\right) |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left(1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{v}\right)^{-\left(\frac{v+p}{2}\right)} ; v > 2 \quad (2.3)$$

$v$  disebut juga parameter bentuk (shape) karena puncak fungsi (2.3) dapat diturunkan atau dinaikkan dengan memberikan variasi nilai  $v$  ini.

Distribusi  $t$  multivariat dapat dipandang sebagai sebuah distribusi normal multivariat rata-rata terboboti dengan bobot  $u$  berdistribusi *Gamma*  $(\frac{1}{2}v, \frac{1}{2}v)$ , dengan fungsi kepadatan peluang sebagai berikut (Meng & Rubin, 1993) :

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, v) = \int_0^\infty f\left(\mathbf{x}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{u}\right) \gamma(u|\alpha, \beta) du \quad (2.4)$$

dengan :

$$f\left(\mathbf{x}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{u}\right) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{\left(\frac{2\pi}{u}\right)^{p/2}} \exp\left(-\frac{u}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \text{ adalah densitas dari } N\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{u}\right)$$

$$\text{dan } \gamma(u|\alpha, \beta) = \frac{\beta^\alpha u^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta u) I(u); I(u) = \begin{cases} 1, u > 0 \\ 0, \text{lainnya} \end{cases}$$

## 2.4 Model Based-Clustering

Metode pengelompokan objek yang banyak dikenal adalah metode *clustering* hirarki (*single linkage*, *complete linkage*, *average linkage* dan *Ward's linkage*) dan *clustering* non hirarki ( *K-means*). Namun metode pengelompokan tersebut tidak mempunyai dasar pengelompokan secara statistik. Adapun metode pengelompokan yang memperhatikan model statistik disebut dengan *model-based clustering* (MBC). Model ini pertama kali digunakan oleh Banfield & Raftery (1993) untuk pengelompokan objek dalam populasi. Asumsi yang digunakan pada *model-based clustering* adalah dalam suatu populasi dapat diambil subpopulasi yang mempunyai distribusi peluang tertentu dan masing-masing subpopulasi mempunyai parameter yang berbeda. Keseluruhan subpopulasi mempunyai distribusi peluang *mixture* dengan proporsi berbeda untuk setiap subpopulasi. Asumsi ini mengarahkan pada model probabilitas matematika yaitu model ***finite mixture***. Saat ini penggunaan model *finite mixture* pada *clustering* telah berkembang sangat cepat dan menjadi salah satu metode *clustering* yang populer.

Banfield dan Raftery (1993) mengembangkan kerangka *model-based clustering* menggunakan dekomposisi *eigenvalue* dari matriks varians kovarians ( $\Sigma_g$ ) sebagai berikut :

$$\Sigma_g = \lambda_g \mathbf{D}_g \mathbf{A}_g \mathbf{D}_g^T \quad (2.5)$$

dengan :

$\lambda_g$  adalah nilai skalar yang menunjukkan volume elips.

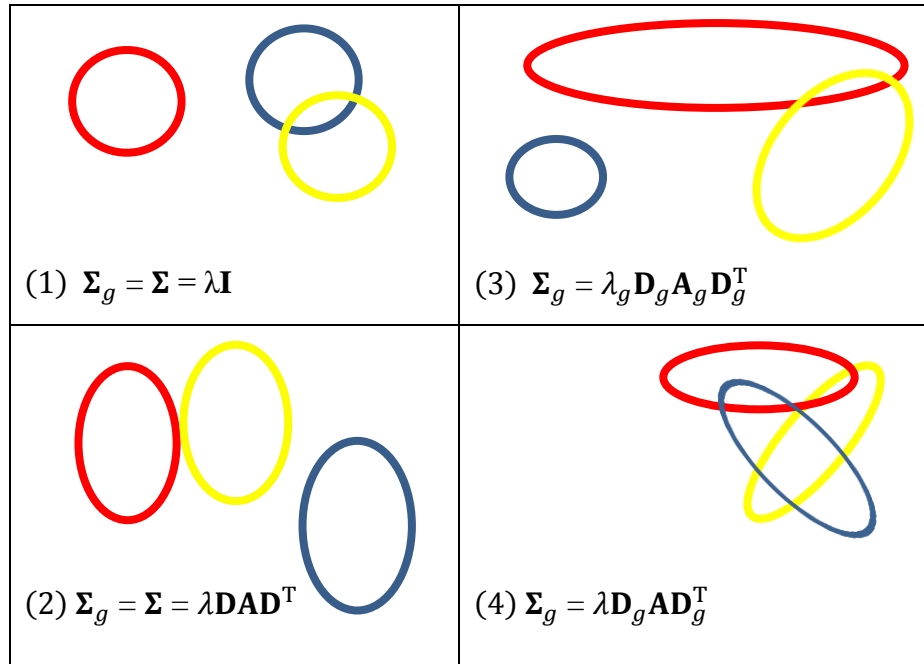
$\mathbf{D}_g$  adalah matriks ortogonal *eigenvector* yang merupakan orientasi dari komponen utama  $\Sigma_g$ .

$\mathbf{A}_g$  adalah diagonal matriks dengan elemen-elemen yang proporsional pada *eigenvalue*  $\Sigma_g$  dan menunjukkan kontur dari fungsi kepadatannya.

Orientasi, volume dan bentuk dari distribusi dapat diestimasi dari data dan mempunyai bentuk bermacam-macam antar *cluster* atau dapat saling memotong antar *cluster*, sebagaimana diilustrasikan pada gambar 2.1.

Keterangan :

1.  $\Sigma_g = \Sigma = \lambda \mathbf{I}$  *cluster* yang terbentuk adalah *spherical* (bola) dan mempunyai volume yang sama besar antar *cluster*.
2.  $\Sigma_g = \Sigma = \lambda \mathbf{DAD}^T$  semua *cluster* yang terbentuk mempunyai bentuk, volume dan orientasi yang sama.
3.  $\Sigma_g = \lambda_g \mathbf{D}_g \mathbf{A}_g \mathbf{D}_g^T$  *cluster* yang terbentuk bisa berbeda pada bentuk, volume dan orientasinya.
4.  $\Sigma_g = \lambda \mathbf{D}_g \mathbf{AD}_g^T$  hanya orientasi dari *cluster* terbentuk yang berbeda.



Gambar 2.1 Contoh Bentuk *Cluster* yang Merepresentasikan Struktur Matriks Varians Kovarians pada *Model-Based Clustering*

Dengan bantuan *software R package teigen*, mampu diidentifikasi 28 model yang mungkin dengan jumlah kelompok maksimal 9 kelompok untuk pengelompokan menggunakan *model-based clustering t* multivariat ( lampiran 1). Sebagai contoh diperoleh model terbaik (nilai ICL terbesar) adalah CUCU dengan jumlah kelompok optimal  $G=2$ . Berdasarkan persamaan (2.5) diketahui model CUCU menunjukkan  $\lambda_g$  berlabel “*C*”*onstrained* yang berarti kedua kelompok memiliki



volume elips yang sama.  $D_g$  berlabel “*U*”*nconstrained* yang berarti orientasi dari kedua kelompok berbeda.  $A_g$  berlabel “*C*”*onstrained* yang berarti kedua kelompok memiliki kontur yang sama.  $D_g^T$  berlabel “*U*”*nconstrained* yang berarti kedua kelompok memiliki derajat bebas yang berbeda. Oleh karena orientasi dan derajat bebas kedua kelompok berbeda, maka kedua kelompok memiliki matrik kovarians ( $\Sigma_g$ ) yang berbeda pula (Andrews, dkk 2011). Model CUCU bersesuaian dengan contoh gambar no (4) pada Gambar 2.1 di atas dengan bentuk cluster yang merepresentasikan struktur matriks varians kovarians  $\Sigma_g = \lambda \mathbf{D}_g \mathbf{A} \mathbf{D}_g^T$ .

#### 2.4.1 Model *Finite Mixture*

Misalkan vektor variabel acak  $\mathbf{x}$  dengan dimensi  $p$  berasal dari distribusi *finite mixture* dengan fungsi kepadatan peluang :

$$f(\mathbf{x}|\boldsymbol{\theta}) = \sum_{g=1}^G \pi_g f_g(\mathbf{x}|\boldsymbol{\theta}_g) \quad (2.6)$$

dengan:

$\boldsymbol{\theta} = (\pi_1, \dots, \pi_G, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G)$  adalah vektor parameter,

$f_g(\mathbf{x}|\boldsymbol{\theta}_g)$  disebut fungsi kepadatan peluang  $\mathbf{x}$  dengan parameter kelompok  $\boldsymbol{\theta}_g$ ,

$G$  adalah jumlah kelompok dan  $\pi_g$  adalah bobot atau proporsi campuran (*mixing proportion*) dari kelompok ke- $g$  dengan batasan :

$$0 \leq \pi_g \leq 1, (g = 1, \dots, G) \text{ dan } \sum_{g=1}^G \pi_g = 1.$$

Tabel 2.1 Struktur Data Kelompok ke- $g$  pada Model *Finite Mixture*

Obyek pengamatan	Variabel						
	1	2	...	$j$	...	$p$	
1	$(x_{11})_g$	$(x_{21})_g$	...	$(x_{j1})_g$	...	$(x_{p1})_g$	$(\mathbf{x}_1)_g$
2	$(x_{12})_g$	$(x_{22})_g$	...	$(x_{j2})_g$	...	$(x_{p2})_g$	$(\mathbf{x}_2)_g$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
$i$	$(x_{1i})_g$	$(x_{2i})_g$	...	$(x_{ji})_g$	...	$(x_{pi})_g$	$(\mathbf{x}_i)_g$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
$n_g$	$(x_{1n_g})_g$	$(x_{2n_g})_g$	...	$(x_{jn_g})_g$	...	$(x_{pn_g})_g$	$(\mathbf{x}_{n_g})_g$

Keterangan:

$(x_{ji})_g$  : Nilai variabel ke -  $j$  obyek pengamatan ke- $i$  kelompok ke-  $g$

$(\mathbf{x}_i)_g$  : Vektor nilai data obyek pengamatan ke- $i$  kelompok ke- $g$ ,

$g$  : Indeks kelompok,  $g = 1, 2, \dots, G$ ,

$j$  : Indeks variabel,  $j = 1, 2, \dots, p$ ,

$i$  : Indeks obyek pengamatan,  $i = 1, 2, \dots, n_g$  ;  $\sum_{g=1}^G n_g = n$ ,

Sedikitnya ada dua proses utama yang perlu dilakukan pada model *finite mixture* yaitu penaksiran parameter dan pemilihan model terbaik yang menggambarkan struktur data. Pembahasan lebih detail tentang model *finite mixture* dan aplikasinya dibahas oleh Titterington, Smith & Makov (1985) dan McLachlan & Peel (2000). Struktur data kelompok ke- $g$  dari sampel acak  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  sebagai nilai obyek pengamatan dari  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  dari sebuah model *finite mixture* adalah sebagai berikut :

#### 2.4.1 Model *Finite Mixture* Normal Multivariat

Sampel acak  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  dianggap berasal dari sebuah model *finite mixture* normal multivariat dengan setiap kelompok ke- $g$  pada model diasumsikan berdistribusi normal multivariat maka fungsi kepadatan probabilitasnya berbentuk:

$$f(\mathbf{x}_i; \pi_g, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) = \sum_{g=1}^G \pi_g f_g(\mathbf{x}_i; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \quad (2.7)$$

dengan  $g = 1, 2, \dots, G$ ,  $i = 1, 2, \dots, n$ ,  $\sum_{g=1}^G \pi_g = 1$ , dan

$$f_g(\mathbf{x}_i; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}_g|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_g) \right\} \quad (2.8)$$

Dan  $\boldsymbol{\mu}_g$  adalah vektor mean dan  $\boldsymbol{\Sigma}_g$  adalah matriks kovarians kelompok ke- $g$  dari distribusi normal multivariat dan  $f_g(\mathbf{x}_i)$  adalah fungsi kepadatan probabilitas kelompok ke- $g$ .

### 2.4.3 Model *Finite Mixture t* Multivariat

Model *finite mixture t* multivariat dianggap sebagai sebuah pendekatan yang lebih *robust* untuk data mengandung *outlier*. Dengan mengasumsikan setiap kelompok pada model *finite mixture* berdistribusi *t* multivariat maka efek adanya *outlier* pada penaksiran parameter model *finite mixture* dapat teratasi (McLachlan, Ng & Bean, 2004). Hal ini karena distribusi *t* memiliki ekor yang lebih panjang dibanding distribusi normal sehingga data pengamatan yang tidak biasa (*atypical*) dari sebuah kelompok akan mendapat bobot yang lebih rendah dalam penghitungan nilai penaksir parameternya (Peel & McLachlan, 2000).

Sampel acak  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  dianggap berasal dari sebuah model *finite mixture t* multivariat dengan setiap kelompok ke- $g$  pada model diasumsikan berdistribusi *t* multivariat maka fungsi kepadatan probabilitasnya berbentuk:

$$f(\mathbf{x}_i; \pi_g, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, v_g) = \sum_{g=1}^G \pi_g f_g(\mathbf{x}_i; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, v_g), \quad (2.9)$$

dengan  $g = 1, 2, \dots, G$ ;  $i = 1, 2, \dots, n$ ,  $\sum_{g=1}^G \pi_g = 1$ , dan

$$f_g(\mathbf{x}_i; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, v_g) = \frac{\Gamma\left(\frac{v_g + p}{2}\right) |\boldsymbol{\Sigma}_g|^{-1/2}}{(\pi v_g)^{\frac{p}{2}} \Gamma\left(\frac{v_g}{2}\right)} \left(1 + \frac{\delta(\mathbf{x}_i, \boldsymbol{\mu}_g; \boldsymbol{\Sigma}_g)}{v_g}\right)^{-\left(\frac{v_g + p}{2}\right)} \quad (2.10)$$

dan

$$\delta(\mathbf{x}_i, \boldsymbol{\mu}_g; \boldsymbol{\Sigma}_g) = (\mathbf{x}_i - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_g) \quad (2.11)$$

adalah jarak Mahalanobis kuadrat antara  $\mathbf{x}_i$  dan  $\boldsymbol{\mu}_g$  (dengan  $\boldsymbol{\Sigma}$  sebagai matrik kovarians). Dalam hal ini  $\boldsymbol{\mu}_g = [\mu_{1g} \ \mu_{2g} \ \dots \ \mu_{pg}]^T$  adalah vektor rata-rata kelompok ke- $g$  dengan  $\mu_{jg} = E[X_{jg}]$  dan

$$\boldsymbol{\Sigma}_g = \begin{bmatrix} \sigma_{11g} & \sigma_{12g} & \dots & \sigma_{1pg} \\ \sigma_{12g} & \sigma_{22g} & \dots & \sigma_{2pg} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1pg} & \sigma_{2pg} & \dots & \sigma_{ppg} \end{bmatrix}$$

adalah matriks varians kovarians pada kelompok ke- $g$  dengan  $\sigma_{jjg} = \text{Var}[X_{jg}]$  dan  $\sigma_{jj^*g} = \text{Cov}[X_{jg}, X_{j^*g}]$ ;  $j \neq j^*$  serta  $v_g$  adalah derajat bebas kelompok ke- $g$ . Ketika  $v_g \rightarrow \infty$  maka pendekatan dengan distribusi *t* ini akan mengarah pada distribusi normal. Sehingga menurut McLachlan & Peel (2000), parameter  $v_g$  dapat dianggap sebagai parameter pengontrol *robust* (*robustness tuning*).

## 2.5 Penaksiran Parameter pada Model *Finite Mixture* dengan Metode *Maximum Likelihood*

Untuk menaksir parameter *finite mixture*, banyak metode telah dikembangkan seperti metode momen, metode *minimum-distance*, *maximum likelihood* (ML) dan pendekatan Bayesian. Pada penelitian ini akan digunakan metode *maximum likelihood* untuk mendapatkan penaksir parameter *model-based clustering*. Metode ML memiliki kelebihan yaitu teknik penaksiran parameternya lebih mudah. Teknik ini hanya dapat digunakan bilamana distribusi populasi diketahui. Selain itu ML sangat sensitif terhadap data ekstrim. Data ekstrim ini sangat berpengaruh terhadap nilai-nilai *mean* ataupun varians. Fungsi likelihood model *finite mixture* pada (2.6) adalah:

$$L_M(\boldsymbol{\vartheta}|\mathbf{x}_i) = \prod_{i=1} \sum_{g=1} \pi_g f_g(\mathbf{x}_i|\boldsymbol{\theta}_g) \quad (2.12)$$

Untuk memudahkan dalam menghitung, seringkali digunakan  $\ln L_M(\boldsymbol{\vartheta}|\mathbf{x}_i)$ . Oleh karena fungsi logaritma merupakan fungsi naik maka fungsi  $\ln$  likelihoodnya adalah sebagai berikut:

$$\ln L_M(\boldsymbol{\vartheta}|\mathbf{x}_i) = \sum_{i=1}^n \ln \sum_{g=1}^G \pi_g f_g(\mathbf{x}_i|\boldsymbol{\theta}_g) \quad (2.13)$$

dengan  $\boldsymbol{\vartheta} = (\pi_1, \pi_2, \dots, \pi_g, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \dots, \boldsymbol{\vartheta}_G)$ .

$\hat{\boldsymbol{\vartheta}}$  adalah *maximum likelihood estimator* (MLE) atau penaksir parameter dari  $\boldsymbol{\vartheta}$  yang didapatkan dengan meminimumkan fungsi  $\ln$  likelihood, yaitu

$$\frac{\partial \ln L(\boldsymbol{\vartheta}|\mathbf{x}_i)}{\partial(\boldsymbol{\vartheta}|\mathbf{x}_i)} = 0.$$

Namun penaksir parameter (MLE) tidak dapat diperoleh secara analitik karena fungsi (2.13) mengandung  $\ln$  penjumlahan distribusi *mixture* dan berbentuk *multimodal*.

Jika data  $\mathbf{x}_i$  memiliki kategori/label kelompok maka MLE bisa diperoleh dalam bentuk eksplisit. Misalkan  $\mathbf{y}_i$  adalah data lengkap berlabel yang merupakan pasangan berurut  $(\mathbf{x}_i, \mathbf{z}_i)$  dimana  $\mathbf{x}_i$  adalah data nilai obyek pengamatan dan  $\mathbf{z}_i$  adalah vektor label kelompok.

$$\mathbf{z}_{ig} = \begin{cases} 1, & \mathbf{x}_i \in \text{kelompok ke } -g, \\ 0, & \text{lainnya} \end{cases} \quad i = 1, 2, \dots, n; g = 1, 2, \dots, G$$

Dikarenakan syarat  $\sum_{g=1}^G z_{ig} = 1$ , maka  $\mathbf{z}_i$  bebas dan identik mengikuti distribusi multinomial dengan peluang  $\pi_1, \pi_2, \dots, \pi_g$  dengan fungsi kepadatan probabilitas  $\mathbf{z}_i$  adalah  $\pi_g$  ketika  $z_{ig} = 1$  yaitu :

$$f_g(\mathbf{z}_i) = \prod_{g=1}^G \pi_g^{z_{ig}} \quad (2.14)$$

Jumlah observasi dalam kelompok  $g$  dapat diperoleh dengan menjumlahkan semua variabel label  $z_{ig}$  untuk semua observasi dalam kelompok  $g$  sehingga :

$$n_g = \sum_{i=1}^n z_{ig} \quad (2.15)$$

$$\text{dan } n = \sum_{g=1}^G n_g. \quad (2.16)$$

Fungsi kepadatan probabilitas bersyarat  $\mathbf{x}_i$  diberikan  $\mathbf{z}_i$  adalah fungsi  $f_g(\mathbf{x}_i; \boldsymbol{\theta}_g)$  ketika  $z_{ig} = 1$  yaitu

$$f(\mathbf{x}_i|\mathbf{z}_i) = \prod_{g=1}^G f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)^{z_{ig}} \quad (2.17)$$

Fungsi kepadatan probabilitas bersama dari data lengkap adalah :

$$\begin{aligned} f(\mathbf{y}_i) &= f(\mathbf{x}_i, \mathbf{z}_i) = f(\mathbf{z}_i)f(\mathbf{x}_i|\mathbf{z}_i) \\ &= \prod_{g=1}^G \pi_g^{z_{ig}} [f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)]^{z_{ig}} \end{aligned} \quad (2.18)$$

Sehingga fungsi likelihood data lengkap untuk  $(\boldsymbol{\vartheta}|\mathbf{y}_i)$  adalah

$$\begin{aligned} L_c(\boldsymbol{\vartheta}|\mathbf{y}_i) &= \prod_{i=1}^n f(\mathbf{y}_i) \\ &= \prod_{i=1}^n \prod_{g=1}^G \pi_g^{z_{ig}} [f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)]^{z_{ig}} \end{aligned} \quad (2.19)$$

Dan fungsi ln likelihood data lengkap untuk  $(\boldsymbol{\vartheta}|\mathbf{y}_i)$  adalah

$$\begin{aligned} \ln L_c(\boldsymbol{\vartheta}|\mathbf{y}_i) &= \ln \left( \prod_{i=1}^n \prod_{g=1}^G \pi_g^{z_{ig}} [f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)]^{z_{ig}} \right) \\ &= \sum_{i=1}^n \sum_{g=1}^G z_{ig} [\ln \pi_g + \ln f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)] \end{aligned} \quad (2.20)$$

Variabel label  $\mathbf{z}_i$  dianggap sebagai parameter yang hilang yang harus diestimasi bersamaan dengan estimasi parameter  $(\boldsymbol{\theta}|\mathbf{y}_i)$ . Dikarenakan data model *finite mixture* dapat dipandang sebagai data tidak lengkap dengan data hilang vektor adalah label  $\mathbf{z}_i$  maka penaksiran parameter dengan metode ML dapat dilakukan dengan algoritma *expectation maximization* (EM) yang diajukan Dempster, Laird & Rubin (1977).

## 2.6 Algoritma EM dan ECM

Proses iterasi mendapatkan penaksir parameter *model-based clustering* dilakukan dengan algoritma EM untuk mempermudah dalam memperoleh penaksir parameter (Dempster, Laird & Rubin, 1977). Data observasi dianggap sebagai data tidak lengkap karena label kelompok belum diketahui. Label kelompok ini disebut sebagai variable laten yang diperlakukan sebagai data hilang dalam algoritma EM. Algoritma EM adalah algoritma optimasi iteratif untuk memaksimumkan fungsi likelihood dari model probabilistik dengan data hilang (*missing data*). Jika label ini diketahui, maka akan didapatkan estimasi parameter di setiap distribusi komponen dengan membagi observasi ke dalam kelompok masing-masing.

Algoritma EM menggunakan iterasi 2 langkah sampai mencapai hasil yang konvergen, yaitu langkah Ekspektasi (*E-step*) dan langkah Maksimisasi (*M-step*).

### 1. Langkah Ekspektasi (*E-step*)

Langkah ini dilakukan dengan menghitung nilai ekspektasi dari fungsi logaritma likelihood data lengkap yaitu

$$\ln L_c(\boldsymbol{\theta}|\mathbf{y}_i) = \sum_{i=1}^n \sum_{g=1}^G z_{ig} [\ln(\pi_g) + \ln(f_g(\mathbf{x}_i|\boldsymbol{\theta}_g))] \quad (2.21)$$

Berdasarkan ekspektasi dari fungsi (2.21) didapatkan penaksir parameter

$$\hat{z}_{ig} = \frac{\hat{\pi}_g f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)}{\sum_{g=1}^G \hat{\pi}_g f_g(\mathbf{x}_i|\boldsymbol{\theta}_g)} \quad (2.22)$$

### 2. Langkah Maksimisasi (*M-step*)

Langkah maksimisasi digunakan untuk memaksimalkan nilai harapan fungsi likelihood persamaan (2.19) terhadap  $\hat{\pi}_g$ ,  $\hat{\mu}_g$  dan  $\hat{\Sigma}_g$  dengan nilai  $\hat{z}_{ig}$  diperoleh dari langkah Ekspektasi (E-step).

$$\hat{z}_g = \frac{n_g}{n} ; \hat{\mu}_g = \frac{\sum_{i=1}^n \hat{z}_{ig} \mathbf{y}_i}{n_g}$$

Perhitungan  $\hat{\Sigma}_g$  menggunakan dekomposisi nilai eigen dibahas pada penelitian Celeux dan Govaert (1995).

Algoritma EM dilakukan terus-menerus sampai mendapatkan hasil iterasi yang konvergen. Setelah Algoritma EM mendapatkan nilai konvergen, anggota *cluster* dikelompokkan menggunakan metode klasifikasi *Maximum a Posteriori* (MAP) sebagai berikut:

$$\text{MAP} \{ \hat{z}_{ig} \} = \begin{cases} 1 & \text{jika } \max \{ \hat{z}_{ig} \} \in ke - g \\ 0 & \text{lainnya} \end{cases} \quad (2.23)$$

Untuk mendapatkan penaksir parameter *model-based clustering mixture t* multivariat dengan metode maksimum likelihood, algoritma EM akan memakan waktu yang cukup lama jika  $v_g$  tidak diketahui. Oleh sebab itu dikembangkan algoritma *Expectation Conditional Maximization* (ECM) untuk mengatasi hal tersebut (McLachan & Peel, 2000).

Langkah-langkah dalam algoritma ECM sama dengan langkah pada algoritma EM. Namun pada algoritma ECM, langkah M diberikan suatu kondisi yang dapat meminimalkan waktu iterasi parameter.  $\boldsymbol{\vartheta}$  dibagi menjadi  $\boldsymbol{\vartheta}_1$  dan  $\vartheta_2$  dimana  $\boldsymbol{\vartheta}_1 = (\pi_1, \dots, \pi_{g-1}, \theta)$  dan  $\vartheta_2 = v$ . Pada iterasi ke  $(k+1)$ , langkah E dalam algoritma ECM sama dengan algoritma EM, tetapi langkah M dalam algoritma EM diganti dengan 2 langkah CM sebagai berikut (McLachlan & Peel, 2000) :

1. Langkah CM-1. Hitung  $\boldsymbol{\vartheta}_1$  iterasi ke  $(k+1)$  dengan memaksimumkan  $\ln L_C(\boldsymbol{\vartheta}_1, \vartheta_2 | \mathbf{y}_i)$  dan menganggap  $\vartheta_2$  *fixed* pada iterasi ke -  $k$  sehingga  $v$  *fixed* pada iterasi ke -  $k$ .
2. Langkah CM-2. Hitung  $\vartheta_2$  iterasi ke  $(k+1)$  dengan memaksimumkan  $\ln L_C(\boldsymbol{\vartheta}_1, \vartheta_2 | \mathbf{y}_i)$  dan  $\boldsymbol{\vartheta}_1$  *fixed* pada iterasi ke -  $(k+1)$ .

Walaupun  $\theta_1$  dan  $\theta_2$  dihitung secara terpisah, akan tetapi hasil yang diperoleh dari langkah CM tersebut ekuivalen dengan langkah M pada algoritma EM. Oleh karena itu tidak ada perbedaan antara Algoritma EM dan ECM (Peel & McLachlan, 2000).

## 2.7 Pemilihan Model Terbaik

Dalam model *finite mixture* untuk *clustering*, pemilihan model terbaik yang menggambarkan struktur data dapat dilakukan melalui pendekatan likelihood dan pendekatan bayesian, pembahasannya dapat dilihat pada McLachlan & Peel (2000). Pemilihan model terbaik berkaitan erat dengan estimasi banyak kelompok  $G$ . Sehingga terjadi *trade-offs* antara pemilihan model terbaik dengan banyak kelompok, dimana jika sebuah model sederhana dipilih maka lebih banyak kelompok dibutuhkan untuk menggambarkan struktur data, namun jika semakin kompleks model yang dipilih maka semakin sedikit kelompok yang memenuhi (Fraley & Raftery, 2002). Ukuran informasi berikut ini adalah ukuran yang akan dipakai sebagai dasar pemilihan model dalam penelitian ini.

### 2.7.1 Integrated Completed Likelihood (ICL)

Pemilihan *model-based clustering* terbaik bisa menggunakan kriteria ICL (Bienarcki, Celeux & Govaert, 2000). Prinsipnya adalah memaksimumkan fungsi likelihood data lengkap  $\mathbf{y}_i = (\mathbf{x}_i, \mathbf{z}_i)$  sehingga didapatkan rumus ICL sebagai berikut :

$$ICL_g = \ln f(\mathbf{y}_i) - \frac{p}{2} \ln(n) \quad (2.24)$$

Dengan  $f(\mathbf{y}_i) = f(\mathbf{x}_i, \mathbf{z}_i)$  adalah fungsi kepadatan peluang bersama data lengkap.  $p$  adalah banyaknya parameter, dan  $n$  adalah banyaknya observasi.

Pada penelitian yang dilakukan oleh Bienarcki dkk (2000) dan Baudry dkk., (2013) didapatkan kesimpulan bahwa ICL mampu memberikan perkiraan jumlah komponen *mixture* yang stabil dan reliabel baik pada data realita maupun data simulasi, dilihat dari sudut pandang pengelompokan. Akan tetapi, ICL dapat



mengabaikan jumlah komponen untuk data simulasi yang timbul dari komponen *mixture* yang tidak terpisah dengan baik (*overlap*). Model terbaik yang dipilih kriteria ICL adalah model dengan nilai ICL paling besar.

### 2.7.2 Minimum Message Length (MML)

Penggunaan MML dalam model *finite mixture* pertama kali diperkenalkan oleh Wallace & Boulton (1968). MML merupakan gabungan metode estimasi titik berdasar prinsip *Invariant Bayesian* dan pemilihan model berdasar teori *Shannon's Information* yang menghubungkan ilmu komputer dan statistika. Dalam kerangka MML yang berdasarkan *inductive inference*, data dianggap dapat membentuk sebuah *message*, suatu informasi data yang dikode menjadi *binary string (message)* yang dikirim dari pengirim ke penerima (secara imajiner). Sehingga prinsip MML adalah mencari sebuah model yang dapat meminimalkan *optimal coding length* dari sebuah *message*. Dengan kata lain MML memilih sebuah model yang cocok dengan data.

Wallace & Freeman (1987) berpendapat estimasi secara statistik dapat dilakukan sebagai proses *coding*. Dalam kasus *clustering*, sebuah *message* terdiri dari 2 bagian dimana bagian pertama menjelaskan *length* ruang parameter pada model yang diberikan dan bagian kedua menjelaskan *length* data pengamatan berdasar model tersebut. Wallace & Freeman mengajukan formula *message length* sebagai berikut:

$$MessLen = -\ln pr(\boldsymbol{\theta}) + \frac{1}{2}\ln|F(\boldsymbol{\theta})| - \ln f(\mathbf{x}|\boldsymbol{\theta}) - \frac{n_p}{2}\ln K_{n_p} + \frac{n_p}{2} \quad (2.25)$$

Dengan  $pr(\boldsymbol{\theta})$  adalah distribusi prior dari nilai parameter,  $F(\boldsymbol{\theta})$  adalah *expected* matriks informasi Fisher,  $f(\mathbf{x}|\boldsymbol{\theta})$  adalah fungsi likelihood model *mixture*,  $n_p$  adalah jumlah parameter yang diestimasi pada model *mixture*, dan  $K_{n_p}$  adalah konstanta *Lattice* pada dimensi  $n_p$ .

Berdasarkan formula *message length* yang diajukan Wallace & Freeman (1987) dan mengasumsikan seluruh parameter berdistribusi *noninformative Jeffreys prior*, Figueiredo & Jain (2002) memilih sebuah model dengan  $G$  kelompok yang memaksimalkan persamaan 2.25.

$$\ln L(\boldsymbol{\vartheta}) = \left[ \frac{m_{\vartheta}}{2} \sum_{g=1}^G \ln \frac{n\pi_g}{12} + \frac{G}{2} \ln \frac{n}{12} + \frac{G(m_{\vartheta} + 1)}{2} \right] \quad (2.26)$$

Dengan penalti MML adalah :

$$\frac{m_{\vartheta}}{2} \sum_{g=1}^G \ln \frac{n\pi_g}{12} + \frac{G}{2} \ln \frac{n}{12} + \frac{G(m_{\vartheta} + 1)}{2} \quad (2.27)$$

Keterangan :

$\ln L(\boldsymbol{\vartheta})$  : Fungsi ln likelihood,

$m_{\vartheta}$  : Banyaknya parameter pada setiap kelompok pada model *mixture*,

$n$  : Banyaknya obyek pengamatan (ukuran sampel),

$\pi_g$  : Bobot (*mixing proportion*) pada kelompok ke- $g$  dan

$G$  : Banyaknya kelompok pada model *mixture*.

Dimana  $m_{\vartheta} = 2p + 1$ , untuk  $p$  = jumlah variabel.

Menurut Figueiredo & Jain (2002), fungsi penalti MML ini dapat diaplikasikan untuk semua jenis model *finite mixture* parametrik.

## 2.8 Deteksi *Outlier* Multivariat

Data tidak bersitribusi normal multivariat atau matriks varians-kovariansnya tidak homogen dapat disebabkan ada observasi yang mempunyai pola berbeda dengan sebagian besar pola data. Observasi tersebut disebut pencilan (*outlier*). Dalam kasus pengelompokan dengan metode *model-based clustering*, *outlier* dapat menjadikan hasil pengelompokan kurang tepat dan penaksir parameter menjadi bias (McLachlan & Peel, 2000). Oleh sebab itu, sangat penting memeriksa keberadaan *outlier*.

Metode yang digunakan untuk memeriksa keberadaaan *outlier* multivariat adalah perhitungan Jarak Mahalanobis yang didefinisikan sebagai berikut :

$$MD_i = [(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})]^{1/2}, i = 1, 2, \dots, n \quad (2.28)$$

dimana  $\bar{\mathbf{x}}$  adalah vektor *mean* sampel dan  $\mathbf{S}$  adalah matriks varians kovarians sampel.

Suatu observasi dikatakan *outlier* jika nilai akar kuadrat jarak Mahalanobis  $> \chi_{p;1-\alpha/2}^2$  dimana  $p$  adalah derajat bebas (Rousseeuw & Van

Zomeran, 1990). Namun jarak mahalanobis sangat sensitif terhadap *outlier*, sehingga dapat merubah nilai  $\bar{\mathbf{x}}$  dan  $\mathbf{S}$  yang mengakibatkan kesalahan identifikasi data *non-outlier* menjadi *outlier*. Untuk mengatasi hal tersebut digunakan jarak *robust* (RD) yang lebih robust terhadap *outlier* (Sunaryo, Setiawan dan Siagian, 2011 dalam (Siagian, 2014)). RD menggunakan estimasi matriks varians kovarians dan *mean* berdasarkan MCD (*Minimum Covariance Determinant*).

Jika diasumsikan  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  adalah sampel acak dari sejumlah  $n$  obyek observasi dalam suatu ruang berdimensi  $R^p$ , maka estimasi MCD bertujuan mendapatkan  $h_i$  subset sampel dari  $n$  obyek observasi dengan ukuran maksimum  $h_i = ((n + p + 1)/2)$  yang memiliki determinan matriks varians kovarians terkecil.

$$MCD = \min\{\det(C)_1, \det(C)_2, \dots, \det(C)_{\binom{n}{h}}\}$$

Jarak Robust (*Robust Distance*) diperoleh dengan rumus sebagai berikut :

$$RD_i = [(\mathbf{x}_i - \mathbf{T})^T \mathbf{C}^{-1} (\mathbf{x}_i - \mathbf{T})]^{1/2}, i = 1, 2, \dots, n \quad (2.29)$$

Dimana  $\mathbf{T}$  adalah penaksir *mean* sampel yang *robust* berdasarkan MCD (rata-rata dari  $h$  subset sampel) :

$$\mathbf{T}_i = \frac{1}{h_i} \sum_{j=1}^{h_i} \mathbf{x}_j, \quad (2.30)$$

dan  $\mathbf{C}$  adalah penaksir matriks kovarians yang *robust* dari  $h$  subset sampel :

$$\mathbf{C}_i = \frac{1}{h_i - 1} \sum_{j=1}^{h_i} (\mathbf{x}_j - \mathbf{T}_i)(\mathbf{x}_j - \mathbf{T}_i)^T. \quad (2.31)$$

Roussew dan Van Zomeran (1990) menggunakan RD dan kuantil dari distribusi  $\chi_p^2$  sebagai nilai *cut off* untuk mendeteksi *outlier* multivariat. Data ke- $i$  dengan nilai  $RD > \sqrt{\chi_{p;1-\alpha/2}^2}$  didefinisikan sebagai *outlier* multivariat.

Pada penelitian ini, deteksi *outlier* menggunakan fungsi *distance-distance plot* (*dd plot*) pada package *mvoutlier software* R.

## 2.9 Skewness dan Kurtosis

Salah satu metode untuk memeriksa distribusi normal multivariat adalah menghitung ukuran skewness dan kurtosis data. Menurut Mardia (1970) pemeriksaan distribusi normal multivariat menggunakan ukuran multivariate skewness ( $\beta_{1,p}$ ) dan kurtosis data ( $\beta_{2,p}$ ) dengan rumus sebagai berikut :

$$\beta_{1,p} = E((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))^3 \text{ dan } \beta_{2,p} = E((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))^2 \quad (2.32)$$

Untuk sampel random, ukuran multivariat skewness ( $b_{1,p}$ ) dan kurtosis ( $b_{2,p}$ ) adalah

$$b_{1,p} = \frac{1}{n^2} \left[ \sum_{i,j=1}^n \{(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})\}^3 \right] \quad (2.33)$$

$$b_{2,p} = \frac{1}{n} \left[ \sum_{i,j=1}^n \{(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})\}^2 \right] \quad (2.34)$$

Dengan  $\mathbf{x}_i = [x_{1i} \ x_{2i} \ \dots \ x_{pi}]^T$ ,  $i=1, 2, \dots, n$ , dengan  $n$  adalah banyaknya pengamatan.  $\bar{\mathbf{x}} = [\bar{x}_1 \ \bar{x}_2 \ \dots \ \bar{x}_p]$  adalah vektor rata-rata sampel dan  $\mathbf{S}$  adalah matriks varians kovarians sampel. Data berdistribusi normal jika  $\beta_{1,p} = 0$  dan  $\beta_{2,p} = p(p+2)$ . Berikut merupakan hipotesis yang akan diuji untuk mengetahui apakah data berasal dari populasi yang berdistribusi normal atau bukan.

1. Uji Skewness

$$H_0 : \beta_{1,p} = 0$$

$$H_1 : \beta_{1,p} \neq 0$$

Dengan statistik uji

$$\frac{n}{6} b_{1,p}^2 \sim \chi_{p(p+1)(p+2)/6}^2$$

2. Uji Kurtosis

$$H_0 : \beta_{2,p} = p(p+2)$$

$$H_1 : \beta_{2,p} \neq p(p+2)$$

Dengan statistik uji

$$\frac{b_{2,p} - \frac{n-1}{n+1} p(p+2)}{\sqrt{\frac{8}{n} p(p+2)}} \sim N(0,1)$$

## 2.10 Uji Manova Satu Arah

Menurut Johnson dan Wichern (2007), dalam suatu analisis seringkali terdapat lebih dari satu populasi yang ingin dibandingkan. Sampel acak dikumpulkan dari  $G$  populasi yang disusun sebagai berikut :

Populasi 1 :  $x_{11}, x_{12}, \dots, x_{1n_1}$

Populasi 2 :  $x_{21}, x_{22}, \dots, x_{2n_2}$

⋮

Populasi  $G$  :  $x_{G1}, x_{G2}, \dots, x_{Gn_g}$

*Multivariate analysis of variance* (manova) digunakan untuk mengetahui apakah vektor rata-rata populasi sama, dan jika tidak sama, komponen rata-rata mana yang berbeda secara nyata.

Tabel 2.2 *Multivariate Analysis of Variance* (Manova)

Sumber Variasi	Derajat Bebas	Matriks Jumlah Kuadrat dan Cross Product
Perlakuan	$g - 1$	$B = \sum_{g=1}^G n_g (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})'$
Sisaan (Residual)	$\sum_{g=1}^G n_g - g$	$W = \sum_{g=1}^G \sum_{i=1}^{n_g} [(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)']$
Total (terkoreksi)	$\sum_{g=1}^G n_g - 1$	$B + W = \sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \bar{\mathbf{x}})(\mathbf{x}_{gi} - \bar{\mathbf{x}})'$

Model manova untuk membandingkan  $G$  vektor rata-rata populasi :

$$\mathbf{x}_{gi} = \boldsymbol{\mu} + \boldsymbol{\tau}_g + \boldsymbol{\varepsilon}_{gi} \quad (2.35)$$

dengan  $i = 1, 2, \dots, n$  dan  $g = 1, 2, \dots, G$ ,  $\boldsymbol{\mu}$  vektor rata-rata keseluruhan dan  $\boldsymbol{\tau}_g$  menyatakan pengaruh perlakuan ke- $g$  dengan  $\sum_{g=1}^G n_g \boldsymbol{\tau}_g = 0$  dimana  $\boldsymbol{\varepsilon}_{gi}$  berdistribusi  $\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ .

Berdasarkan model di atas, setiap komponen vektor observasi  $\mathbf{x}_{ij}$  memenuhi model univariat :

$$\mathbf{x}_{gi} = \boldsymbol{\mu}(\text{overall mean}) + \boldsymbol{\tau}_g(\text{treatment effect}) + \boldsymbol{\varepsilon}_{gi}(\text{random error}), \quad (2.36)$$

Dengan  $\mathbf{x}_{gi} \sim N(\boldsymbol{\mu} + \boldsymbol{\tau}_g, \boldsymbol{\Sigma}_g)$  dan  $\boldsymbol{\varepsilon}_{gi} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . Vektor observasi didekomposisi sebagai berikut :

$$\mathbf{x}_{gi} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_g - \bar{\mathbf{x}}) + (\mathbf{x}_{gi} - \bar{\mathbf{x}}_g) \quad (2.37)$$

(observasi) = (rata-rata sampel keseluruhan  $\hat{\boldsymbol{\mu}}$ ) + (pengaruh perlakuan taksiran  $\hat{\boldsymbol{\tau}}$ ) + (sisaan  $\boldsymbol{\varepsilon}_{gi}$ ).

Penjumlahan hasil kali silang terhadap  $g$  dan  $i$  adalah :

$$\begin{aligned} \sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \bar{\mathbf{x}})(\mathbf{x}_{gi} - \bar{\mathbf{x}}) &= \sum_{g=1}^G \sum_{i=1}^{n_g} [(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g) + (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})] [(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g) + (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})] \\ &= \sum_{g=1}^G \sum_{i=1}^{n_g} [(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)' + (\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})' \\ &\quad + (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)' + (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})'] \end{aligned} \quad (2.38)$$

Karena  $\sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \bar{\mathbf{x}}) = \mathbf{0}$  maka

$$\begin{aligned} \sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \bar{\mathbf{x}})(\mathbf{x}_{gi} - \bar{\mathbf{x}}) &= \sum_{g=1}^G \sum_{i=1}^{n_g} (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})' \\ &\quad + \sum_{g=1}^G \sum_{i=1}^{n_g} (\bar{\mathbf{x}}_{gi} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{gi} - \bar{\mathbf{x}})' \\ &= \sum_{g=1}^G n_g (\bar{\mathbf{x}}_g - \bar{\mathbf{x}})(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})' \\ &\quad + \sum_{g=1}^G \sum_{i=1}^{n_g} (\bar{\mathbf{x}}_{gi} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{gi} - \bar{\mathbf{x}})' \end{aligned} \quad (2.39)$$

SSCP Total (terkoreksi) = SSCP Perlakuan (*between*) + SSCP Residual (*within*)  
dimana matriks SSCP Residual dapat dinyatakan sebagai :

$$\begin{aligned} \mathbf{W} &= \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' \\ &= (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \cdots + (n_g - 1)\mathbf{S}_g \end{aligned} \quad (2.40)$$

SSCP = *Sum of Squares and Cross Products* (Jumlah Kuadrat dan hasil Kali

Silang). Hipotesisnya adalah sebagai berikut :

$H_0: \tau_1 = \tau_2 = \dots = \tau_G = \mathbf{0}$  (tidak ada perbedaan antar kelompok)

Kriteria pengujian (Statistik uji jika  $H_0$  benar) yang digunakan sebagai berikut

(Khattree, R., & Naik, D. N, 2000) :

1. Kriteria Wilk's Lambda

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \prod_{g=1}^G \left( \frac{1}{1 + \hat{\lambda}_g} \right)$$

Dengan  $\hat{\lambda}_1, \dots, \hat{\lambda}_g$  adalah nilai eigen dari  $\mathbf{W}^{-1}\mathbf{H}$

2. Kriteria Lawley Hotelling (Hotelling-Lawley Trace)

$$\mathbf{U} = tr(\mathbf{W}^{-1}\mathbf{B}) = \sum_{g=1}^G \lambda_{gt}$$

3. Kriteria Trace Pillai (Pillai's Trace)

$$\mathbf{V} = tr[(\mathbf{W} + \mathbf{B})^{-1}\mathbf{W}] = \sum_{g=1}^G \frac{\lambda_g}{1 + \lambda_g}$$

4. Kriteria Akar Maksimum Roy (Roy's Greatest Root)

$$\lambda_{max} = \max(\lambda_1, \lambda_2, \dots, \lambda_g)$$

Pendekatan F untuk kriteria pengujian tersebut dirangkum dalam tabel berikut.

Tabel 2.3 Pendekatan F untuk kriteria pengujian Manova

Kriteria	F <sub>hitung</sub>	Distribusi pendekatan F di bawah $H_0$
Wilks' Lambda	$\frac{qt - 2u}{rr_t} \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}}$	$F_{(rr_t, qt-2u)}$
Lawley Hotelling	$\frac{2(sm_2 + 1)}{s^2(2m_1 + s + 1)} \mathbf{U}$	$F_{(s(2m_1+s+1), 2(sm_2+1))}$
Trace Pillai	$\frac{2m_2 + s + 1}{2m_1 + s + 1} \frac{\mathbf{V}}{\mathbf{S} + \mathbf{V}}$	$F_{(s(2m_1+s+1), s(2m_2+s+1))}$
Akar Maksimum Roy	$\frac{n - g - h + r - 1}{h} \lambda_{max}$	$F_{(h, n-g-h+r-1)}$

dengan

$$\begin{aligned}
r_t &= \text{rank matriks } (\mathbf{W} + \mathbf{B}) & m_1 &= [|r - r_t| - 1]/2 \\
r &= \text{banyaknya variabel respon} & m_2 &= (n - g - r_t - 2)/2 \\
s &= \min(r_t, r) & q &= (n - g = 1) - \frac{(r_t - r + 1)}{2} \\
h &= \max(r_t, r) & u &= (rr_t - 2)/4 \\
t &= \begin{cases} \sqrt{r^2 r_t^2 - \frac{4}{(r_t^2 + r^2 - 5)}} & , \text{ jika } r_t^2 + r^2 - 5 > 0 \\ 1, & \text{ lainnya} \end{cases}
\end{aligned}$$

Selain Wilk's Lambda, statistik uji lain yang lebih *robust* untuk data tidak berdistribusi normal multivariat adalah statistik Pillai's Trace (Rencher, 2002).

## 2.11 Survei Angkatan Kerja Nasional (Sakernas)

Data ketenagakerjaan yang dikumpulkan oleh Badan Pusat Statistik (BPS) melalui sensus dan survei antara lain : Sensus Penduduk (SP), Survei Penduduk Antar Sensus (SUPAS), Survei Sosial Ekonomi Nasional (SUSENAS) dan Survei Angkatan Kerja Nasional (SAKERNAS). Dari survei-survei tersebut, hanya Sakernas yang dirancang khusus untuk mengumpulkan data yang dapat menggambarkan keadaan umum ketenagakerjaan antar periode pencacahan (BPS, 2014). Kegiatan pengumpulan data ketenagakerjaan pertama kali dilaksanakan tahun 1976. Sampai dengan saat ini, Sakernas mengalami berbagai perubahan baik dalam periode pencacahan maupun cakupan sampel wilayah dan rumah tangga. Tahun 1986 sampai dengan 1993 Sakernas dilaksanakan secara triwulanan, tahun 1994 sampai dengan 2001 secara tahunan setiap bulan Agustus, sedangkan tahun 2002 sampai dengan 2004 selain secara tahunan juga dilaksanakan secara triwulanan. Mulai tahun 2005 sampai dengan tahun 2010 Sakernas dilakukan secara semesteran. Dengan semakin mendesaknya tuntutan data ketenagakerjaan baik variasi, kontinuitas, kemutakhiran dan peningkatan akurasi data yang dihasilkan, maka pengumpulan data Sakernas sejak tahun 2011 mulai dilakukan kembali secara triwulanan yaitu; bulan Februari (Triwulan I),



Mei (Triwulan II), Agustus (Triwulan III) dan November (Triwulan IV) yang penyajian datanya dirancang sampai tingkat provinsi. Sakernas Triwulanan ini dimaksudkan untuk memantau indikator ketenagakerjaan secara dini di Indonesia, yang mengacu pada KILM (*The Key Indicators of the Labour Market*) yang direkomendasikan oleh ILO (*The International Labour Organization*). Hasil Sakernas Triwulan I, II, dan IV disajikan sampai tingkat provinsi (jumlah sampel 50.000 rumah tangga). Sementara Sakernas Triwulan III, disajikan sampai tingkat kabupaten/kota, karena jumlah sampel cukup besar sekitar 200.000 rumah tangga, dimana jumlah tersebut terdiri dari 50.000 rumah tangga merupakan sampel Sakernas Triwulanan dan 150.000 rumah tangga sampel Sakernas tambahan. Mulai tahun 2015, Sakernas dilaksanakan 2 kali dalam 1 tahun (semesteran), yaitu dibulan Agustus ditujukan untuk estimasi kabupaten/kota, dan bulan Februari untuk estimasi provinsi.

#### 2.11.1 Kerangka Sampel

Kerangka sampel yang digunakan terdiri dari tiga jenis, yaitu kerangka sampel untuk penarikan sampel tahap pertama, kerangka sampel untuk penarikan sampel tahap kedua dan kerangka sampel untuk penarikan sampel tahap ketiga. Blok sensus dalam kerangka sampel dipilah menjadi dua kelompok, yaitu blok sensus terpilih untuk estimasi tingkat provinsi, dan blok sensus komplemen (sebagai tambahan) untuk estimasi kabupaten.

- Kerangka sampel pemilihan **tahap pertama** adalah daftar wilayah pencacahan (wilcah) SP2010 yang terpilih Susenas Triwulan I yang disertai dengan informasi banyaknya rumah tangga hasil listing SP2010 (Daftar RBL1), muatan blok sensus dominan (pemukiman biasa, pemukiman mewah, pemukiman kumuh), informasi daerah sulit/tidak sulit, dan klasifikasi desa/kelurahan (*rural/urban*).
- Kerangka sampel pemilihan **tahap kedua** adalah daftar blok sensus pada setiap wilcah terpilih.
- Kerangka sampel pemilihan **tahap ketiga** adalah daftar rumah tangga biasa tidak termasuk *institutional household* (panti asuhan, barak

polisi/militer, penjara, dsb) dalam setiap blok sensus sampel hasil pencacahan lengkap SP2010 (SP2010.C1) yang telah dimutakhirkan pada setiap menjelang pelaksanaan survei.

### 2.11.2 Desain Sampel

Pemilihan sampel rumah tangga dirancang dengan penarikan sampel tiga tahap, dengan tahapan sebagai berikut:

**Tahap pertama:** dari daftar wilcah SP2010 dipilih 30.000 wilcah untuk Susenas secara *Probability Proportional to Size* (pps) dengan *size* jumlah rumah tangga SP2010. Kemudian 30.000 wilcah tersebut dialokasikan sama ke dalam empat triwulan, masing-masing sebesar 7.500 wilcah. Dari 7.500 wilcah Susenas Triwulan I, dipilih 5.000 wilcah secara sistematis untuk Sakernas 2011 Triwulan I dan akan digunakan lagi untuk Triwulan II, III, dan IV.

**Tahap kedua:** memilih dua blok sensus pada setiap wilcah terpilih Susenas yang juga terpilih Sakernas secara pps sistematis dengan *size* jumlah rumah tangga SP2010-C1. Selanjutnya blok-blok sensus terpilih dialokasikan secara acak untuk Susenas dan Sakernas. Blok-blok sensus terpilih Sakernas ini digunakan untuk estimasi provinsi dan dibagi ke dalam 4 paket sampel.

Khusus untuk Sakernas Triwulan III, yang diperuntukkan untuk **estimasi kabupaten**, diperlukan tambahan sampel blok sensus. Dari 15.000 sampel wilcah terpilih Susenas Triwulan II dan III masing-masing dipilih 2 blok sensus, satu untuk keperluan Susenas dan yang lainnya untuk Sakernas. Blok sensus untuk Sakernas yang terpilih dari *PSU* Susenas Triwulan II dan III ini selanjutnya digunakan sebagai sampel **blok sensus komplemen** yang merupakan tambahan sampel yang apabila digabungkan dengan blok sensus estimasi provinsi (Sakernas Triwulan III) dapat digunakan untuk estimasi kabupaten.

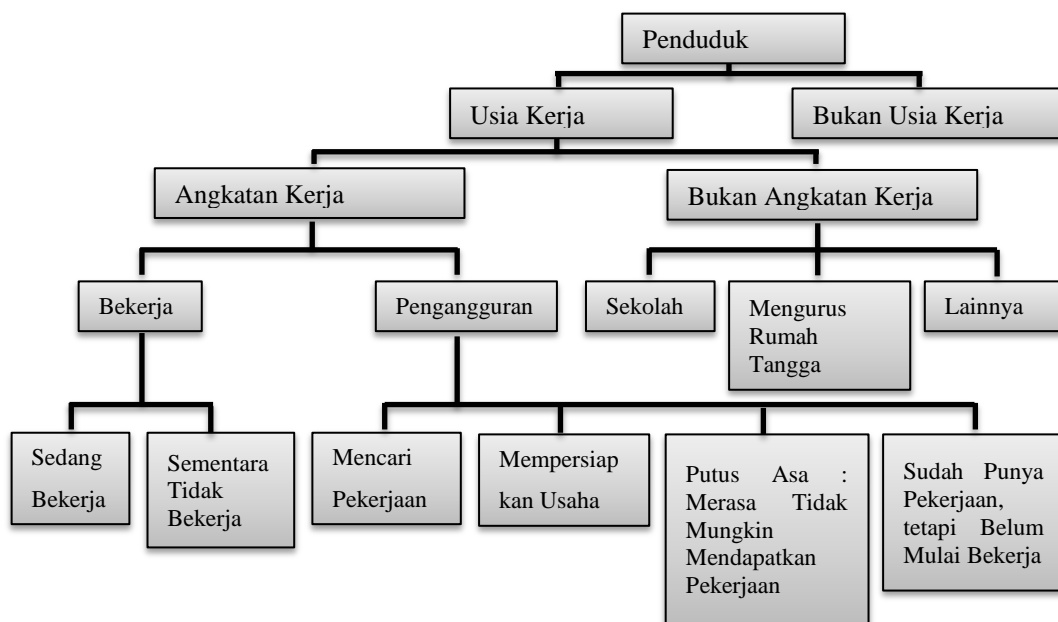
**Tahap ketiga:** memilih 10 rumahtangga secara sistematis berdasarkan hasil pemutakhiran rumah tangga SP2010-C1.

Pemutakhiran frame rumah tangga dilakukan pada setiap awal periode pencacahan sehingga bila terjadi penambahan populasi rumah tangga secara *significance* pada blok sensus terpilih mengharuskan adanya penambahan sampel pada blok sensus tersebut. Akan tetapi apabila terjadi penurunan populasi rumah tangga di suatu

blok sensus tidak serta merta menyebabkan pengurangan sampel rumah tangga **kecuali** sampel rumah tangga yang telah ditentukan karena sesuatu dan lain hal sehingga tidak bisa dicacah kembali.

### 2.11.3 Konsep Umum Ketenagakerjaan

Konsep definisi ketenagakerjaan pada penelitian ini menggunakan konsep BPS yang merujuk pada rekomendasi ILO, seperti tercantum dalam buku “*Surveys of Economically Active Population, Employment, Unemployment and Under employment: An ILO Manual on Concepts and Methods*”, ILO 1992. Oleh karena itu data ketenagakerjaan yang dihasilkan dari berbagai survei di Indonesia dapat dibandingkan secara internasional, tanpa mengesampingkan kondisi ketenagakerjaan spesifik Indonesia.



Gambar 2.2 Bagan Konsep Dasar Tenaga Kerja

Gambar di atas merinci Konsep Dasar Angkatan Kerja (*Standard Labor Force Concept*) pada survei angkatan kerja nasional (Sakernas).

Berikut beberapa konsep umum yang digunakan dalam indikator pasar tenaga kerja.

- Penduduk adalah semua orang yang berdomisili di wilayah geografis Republik Indonesia selama enam bulan atau lebih dan atau mereka yang

berdomisili kurang dari 6 bulan tetapi bertujuan untuk menetap. Berdasarkan konsep dasar ketenagakerjaan, penduduk dikelompokkan menjadi penduduk usia kerja dan penduduk bukan usia kerja. Penduduk usia kerja dibedakan atas dua kelompok, yaitu angkatan kerja dan bukan angkatan kerja. Pengukurannya didasarkan pada periode rujukan (*time reference*), yaitu kegiatan yang dilakukan selama seminggu yang lalu sampai sehari sebelum pencacahan.

- Penduduk usia kerja adalah penduduk berumur 15 tahun dan lebih. Usia kerja merupakan batas usia yang ditetapkan untuk pengelompokan penduduk yang aktif secara ekonomi (*economically active population*). Indonesia menggunakan batas bawah usia kerja 15 tahun, meskipun dalam Sakernas dikumpulkan informasi mulai dari usia 10 tahun dan tanpa batas atas usia kerja. Di negara lain, penentuan batas bawah dan batas atas usia kerja bervariasi sesuai dengan kebutuhan/situasi masing-masing negara.
- Angkatan kerja adalah penduduk usia kerja yang terdiri dari penduduk yang bekerja, pengangguran dan penduduk yang punya pekerjaan namun sementara tidak bekerja.
- Bukan angkatan kerja adalah penduduk usia kerja yang pada periode rujukan survei tidak mempunyai/ melakukan aktivitas ekonomi, baik karena sekolah, mengurus rumah tangga atau lainnya (olahraga, kursus, piknik serta kegiatan sosial seperti berorganisasi dan kerja bakti).
- Bekerja merupakan kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus) dalam seminggu yang lalu. Kegiatan tersebut termasuk kegiatan pekerja tak dibayar yang membantu dalam suatu usaha/kegiatan ekonomi.
- Pasar merupakan tempat perjumpaan antara pembeli dan penjual, di mana barang/jasa atau produk dipertukarkan antara pembeli dan penjual. Ukuran kerelaan dalam pertukaran tersebut biasanya akan muncul suatu tingkat harga atas barang dan jasa yang dipertukarkan tersebut. Ketika seperangkat peraturan dan tata cara yang resmi dijadikan panduan dan batasan hubungan

antara pekerja dan perusahaan, maka saat itulah pasar tenaga kerja tercipta (Ehrenberg dan Smith, (2012).

- Tenaga kerja adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan barang atau jasa baik untuk memenuhi kebutuhan sendiri maupun untuk masyarakat (UU No. 13 tahun 2003 Bab I pasal 1 ayat 2).

Pemilihan sampel rumah tangga Sakernas melalui proses yang tidak singkat. Sebelum dilakukan pemilihan sampel rumah tangga, terlebih dahulu dilakukan pembentukan paket sampel blok sensus dan pembentukan kelompok sampel rumah tangga. Semua proses sampling ini dilakukan untuk menjamin agar rumah tangga yang terpilih memang representatif sebagai sampel untuk estimasi baik tingkat provinsi maupun tingkat kabupaten/kota.

## **2.12 Indikator Pasar Tenaga Kerja (*Key Indicators of The Labor Market KILM*)**

ILO merilis KILM pada tahun 1999 untuk melengkapi program pengumpulan data secara rutin dan untuk meningkatkan penyebaran data pada elemen kunci dari pasar tenaga kerja dunia. Terdapat 20 (dua puluh) indikator yang disusun oleh ILO dan dikelompokkan ke dalam 8 (delapan) kelompok, yaitu:

- 1) Partisipasi di dunia kerja, yang terdiri dari KILM 1, yaitu Tingkat Partisipasi Angkatan Kerja;
- 2) Indikator penduduk yang bekerja, terdiri dari KILM 2 (Rasio Penduduk yang Bekerja Terhadap Jumlah Penduduk), KILM 3 (Penduduk yang Bekerja Menurut Status Pekerjaan Utama), KILM 4 (Penduduk yang Bekerja Menurut Lapangan Usaha), KILM 5 (Pekerja Paruh Waktu), KILM 6 (Penduduk yang Bekerja Menurut Jam Kerja), dan KILM 7 (Penduduk yang Bekerja di Sektor Informal);
- 3) Indikator pengangguran, *underemployment* (setengah penganggur), dan ketidakaktifan, yang terdiri dari KILM 8 (Pengangguran), KILM 9 (Pengangguran pada Kelompok Umur Muda), KILM 10 (Pengangguran Jangka Panjang), KILM 11 (Pengangguran dan Pendidikan), KILM 12

(Setengah Penganggur/*underemployment*), dan KILM 13 (Tingkat Ketidakaktifan);

- 4) Indikator pendidikan dan melek huruf, yang terdiri dari KILM 14 (Pencapaian Pendidikan dan Melek Huruf);
- 5) Indikator upah dan biaya tenaga kerja, yang terdiri dari KILM 15 (Indeks Upah Sektor Manufaktur), KILM 16 (Indikator Upah dan Pendapatan Berdasarkan Jabatan), dan KILM 17 (Upah per Jam);
- 6) Produktivitas Tenaga Kerja yang termuat dalam KILM 18 (Produktivitas Tenaga Kerja);
- 7) Indikator elastisitas tenaga kerja yang termuat dalam KILM 19 (Elastisitas Tenaga Kerja);
- 8) Indikator kemiskinan, pekerja miskin, dan distribusi pendapatan yang tertuang dalam KILM 20 (Indikator Kemiskinan, Penduduk Bekerja yang Miskin, dan Distribusi Pendapatan).

Berikut penjelasan singkat 10 indikator KILM beserta variabel-variabel ketenagakerjaan yang digunakan dalam penelitian ini.

#### **2.12.1 KILM 1. Tingkat Partisipasi Angkatan Kerja (TPAK)**

Tingkat Partisipasi Angkatan Kerja (TPAK) adalah ukuran proporsi penduduk umur kerja yang terlibat aktif di pasar tenaga kerja, baik dengan bekerja atau mencari pekerjaan, yang memberikan indikasi ukuran relatif dari pasokan tenaga kerja yang tersedia untuk terlibat dalam produksi barang dan jasa. Rincian angkatan kerja menurut jenis kelamin dan kelompok umur memberikan profil distribusi penduduk yang aktif secara ekonomi.

$$\begin{aligned} TPAK &= \frac{\text{Penduduk Bekerja} + \text{Pengangguran}}{\text{Penduduk Umur Kerja}} \times 100\% \\ &= \frac{\text{Penduduk Angkatan Kerja}}{\text{Penduduk Umur Kerja}} \times 100\% \end{aligned} \quad (2.41)$$

Secara umum, kegunaan indikator ini adalah untuk mengindikasikan besarnya penduduk umur kerja (15 tahun ke atas) yang aktif secara ekonomi di suatu negara atau wilayah, dan menunjukkan besaran relatif dan pasokan tenaga kerja (*labour supply*) yang tersedia untuk produksi barang dan jasa dalam suatu

perekonomian. TPAK diukur sebagai persentase jumlah angkatan kerja terhadap jumlah penduduk umur kerja.

### **2.12.2 KILM 2.Rasio Penduduk yang Bekerja Terhadap Jumlah Penduduk (*Employment to Population Ratio-EPR*)**

EPR didefinisikan sebagai proporsi penduduk umur kerja suatu negara yang berstatus bekerja terhadap penduduk umur kerja. Rasio yang tinggi berarti sebagian besar penduduk suatu negara adalah bekerja, sementara rasio rendah berarti bahwa sebagian besar penduduk tidak terlibat langsung dalam kegiatan yang berhubungan dengan pasar, karena mereka menganggur atau (lebih mungkin) tidak termasuk dalam angkatan kerja, yang sering disebut sebagai Bukan Angkatan Kerja (BAK).

$$EPR = \frac{\text{Penduduk Bekerja}}{\text{Penduduk Umur Kerja}} \times 100\% \quad (2.42)$$

Rasio ini memberikan informasi tentang kemampuan ekonomi untuk menciptakan lapangan kerja. Di banyak negara, indikator ini menghasilkan analisis yang lebih mendalam dibandingkan dengan tingkat pengangguran. Meskipun secara keseluruhan rasio tinggi biasanya dianggap sebagai positif, indikator ini saja tidak cukup untuk menilai tingkat pekerjaan yang layak atau tingkat defisit pekerjaan yang layak. Indikator tambahan diperlukan untuk menilai isu-isu seperti upah/gaji, jam kerja, lapangan kerja sektor informal, setengah pengangguran, dan kondisi kerja. Bahkan, nilai rasio ini bisa tinggi untuk alasan yang tidak selalu positif misalnya, pilihan pendidikan yang terbatas sehingga kaum muda mengambil pekerjaan yang tersedia daripada tinggal di sekolah untuk membangun sumber daya mereka (memilih bekerja demi melanjutkan sekolah/membiayai sekolah). Untuk alasan ini, sangat disarankan bahwa indikator ini harus ditinjau ulang secara kolektif dalam setiap evaluasi kebijakan tenaga kerja di suatu negara/wilayah.

### 2.12.3 KILM 3. Penduduk yang Bekerja Menurut Status Pekerjaan Utama

**Status pekerjaan** adalah jenis kedudukan seseorang dalam melakukan pekerjaan di suatu unit usaha/kegiatan. Pada penelitian ini, status pekerjaan dibedakan menjadi 6 kategori yaitu:

- a. **Karyawan/ Pegawai/ Buruh**, adalah seseorang yang bekerja pada orang lain atau instansi/kantor/perusahaan secara tetap dengan menerima upah/gaji baik berupa uang maupun barang. Buruh yang tidak mempunyai majikan tetap, tidak digolongkan sebagai buruh/karyawan, tetapi sebagai **pekerja bebas**. Seseorang dianggap memiliki majikan tetap jika memiliki 1 (satu) majikan (orang/rumah tangga) yang sama dalam sebulan terakhir, khusus pada sektor bangunan batasannya tiga bulan. Apabila majikannya instansi/lembaga, boleh lebih dari satu majikan.
- b. **Pengusaha/Berusaha dibantu buruh tetap/buruh dibayar**, adalah berusaha atas resiko sendiri dan mempekerjakan paling sedikit satu orang buruh/pekerja tetap yang dibayar.
- c. **Berusaha sendiri dan Berusaha dibantu buruh tidak tetap/buruh tak dibayar**. Berusaha sendiri adalah bekerja atau berusaha dengan menanggung resiko secara ekonomis, yaitu dengan tidak kembalinya ongkos produksi yang telah dikeluarkan dalam rangka usahanya tersebut, serta tidak menggunakan pekerja dibayar maupun pekerja tak dibayar, termasuk yang sifat pekerjaannya memerlukan teknologi atau keahlian khusus.  
Berusaha dibantu buruh tidak tetap/buruh tak dibayar adalah bekerja atau berusaha atas resiko sendiri, dan menggunakan buruh/pekerja tak dibayar dan atau buruh/pekerja tidak tetap.
- d. **Pekerja bebas**, adalah seseorang yang bekerja pada orang lain/majikan/institusi yang tidak tetap (lebih dari 1 majikan dalam sebulan terakhir) di usaha pertanian maupun non pertanian, baik berupa usaha rumah tangga maupun bukan usaha rumah tangga, atas dasar balas jasa dengan menerima upah atau imbalan baik berupa uang maupun barang, dan baik dengan sistem pembayaran harian maupun borongan.

**Majikan** adalah orang atau pihak yang memberikan pekerjaan dengan pembayaran yang disepakati.



- e. **Pekerja keluarga/tak dibayar** adalah seseorang yang bekerja membantu orang lain yang berusaha dengan tidak mendapat upah/gaji, baik berupa uang maupun barang. Pekerja tak dibayar tersebut bisa sebagai istri, anak keponakan, paman, bibi dan teman.
- f. **Pekerja rentan** (*vulnerable employment*) adalah pekerja yang mencakup pekerja dengan status berusaha sendiri, berusaha dibantu buruh tidak tetap/tak dibayar, pekerja bebas dan pekerja keluarga.

#### 2.12.4 KILM 4. Penduduk yang Bekerja Menurut Sektor (Lapangan Usaha)

**Lapangan usaha (sektor)** adalah bidang kegiatan dari pekerjaan/ usaha/ perusahaan/kantor tempat seseorang bekerja. Klasifikasi baku yang digunakan dalam penggolongan lapangan pekerjaan/lapangan usaha adalah Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) 2009 dengan kategori sebagai berikut:

- 1) Pertanian, kehutanan, perburuan, dan perikanan;
- 2) Pertambangan dan penggalian;
- 3) Industri pengolahan;
- 4) Listrik, gas, dan air;
- 5) Bangunan;
- 6) Perdagangan besar, eceran, rumah makan, dan hotel;
- 7) Transportasi, pergudangan, dan komuni-kasi;
- 8) Keuangan, asuransi, usaha persewaan bangunan, tanah, dan jasa perusahaan;
- 9) Jasa kemasyarakatan.

Lapangan usaha pada publikasi KILM dan digunakan pada penelitian ini dibagi menjadi tiga yaitu:

- 1. Sektor Pertanian (*Agriculture*), terdiri dari sektor (1) KBLI 2009.
- 2. Sektor Manufaktur (*Manufacture*), terdiri dari sektor (2), (3), (4) dan (5).
- 3. Sektor Jasa-jasa (*Services*), terdiri dari sektor (6). (7), (8) dan (9).

Pembagian tiga sektor di atas berdasarkan definisi sektor *International Standard Industrial Classification (ISIC) System* (Revisi 2 dan Revisi 3).

#### **2.12.5 KILM 5. Pekerja Paruh Waktu**

Pekerja paruh waktu adalah mereka yang bekerja di bawah jam kerja penuh-waktu normal, tetapi tidak mencari pekerjaan atau tidak bersedia menerima pekerjaan lain. Karena tidak ada definisi yang disepakati secara internasional untuk jumlah minimum jam dalam seminggu yang merupakan pekerjaan penuh-waktu (*full time*), biasanya digunakan batas 35 jam seminggu sebagai jam kerja normal. Beberapa negara yang menggunakan batas 35 jam adalah Republik Korea, Amerika, dan El Salvador .

#### **2.12.6 KILM 6. Penduduk yang Bekerja Menurut Jam Kerja**

KILM ini bertujuan untuk menunjukkan jumlah orang yang dipekerjakan menurut jam bekerja (biasanya atau sebenarnya): kurang dari 25 jam kerja per minggu; antara 25 dan 34 jam; antara 35 dan 39 jam; antara 40 dan 48 jam; antara 49 dan 59 jam; dan 60 jam ke atas.

#### **2.12.7 KILM 7. Penduduk yang Bekerja di Sektor Informal**

Konferensi Internasional Statistik Tenaga Kerja (*The International Conference of Labour Statisticians-ICLS*) ke-15 mendefinisikan sektor informal sebagai unit produksi dalam usaha rumah tangga yang dimiliki oleh rumah tangga. Mereka yang bekerja di sektor informal terdiri semua orang yang selama periode acuan tertentu bekerja pada setidaknya satu unit produksi yang memenuhi konsep sektor informal, terlepas dari status mereka dalam pekerjaan itu apakah pekerjaan utama atau pekerjaan sekunder. Resolusi ICLS memperbolehkan beberapa variasi konsep nasional. Akibatnya, informasi untuk indikator ini sering didasarkan pada definisi nasional dan pengukuran ekonomi informal. Selain itu, pekerjaan informal dicirikan oleh ketiadaan kontrak, perlindungan sosial, hak untuk berbagai jaminan dan tidak tunduk pada undang-undang tenaga kerja dan pendapatan pajak.

#### **2.12.8 KILM 8. Pengangguran**

Penganggur adalah mereka yang tidak mempunyai pekerjaan, sedang mencari pekerjaan, dan bersedia untuk bekerja. Tingkat Pengangguran Terbuka

(TPT) merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja.

$$TPT = \frac{\text{Jumlah Pengangguran}}{\text{Jumlah Penduduk Bekerja} + \text{Pengangguran}} \times 100\% \quad (2.43)$$
$$= \frac{\text{Jumlah Pengangguran}}{\text{Jumlah Angkatan Kerja}} \times 100\%$$

**Penganggur terbuka**, terdiri dari:

- a. Mereka yang tak punya pekerjaan dan mencari pekerjaan.

Mencari pekerjaan adalah kegiatan seseorang yang pada saat survei, orang tersebut sedang mencari pekerjaan. Seperti mereka:

- Yang belum pernah bekerja dan sedang berusaha mendapatkan pekerjaan.
- Yang sudah pernah bekerja, karena sesuatu hal berhenti atau diberhentikan dan sedang berusaha untuk mendapatkan pekerjaan.
- Yang bekerja atau mempunyai pekerjaan, tetapi karena sesuatu hal masih berusaha untuk mendapatkan pekerjaan lain.

- b. Mereka yang tak punya pekerjaan dan mempersiapkan usaha.

Mempersiapkan suatu usaha adalah suatu kegiatan yang dilakukan seseorang dalam rangka mempersiapkan suatu usaha/pekerjaan yang "baru", yang bertujuan untuk memperoleh penghasilan/keuntungan atas resiko sendiri, baik dengan atau tanpa mempekerjakan buruh/pekerja dibayar maupun tidak dibayar. Mempersiapkan yang dimaksud adalah apabila "tindakannya nyata" •, seperti: mengumpulkan modal atau perlengkapan/alat, mencari lokasi/tempat, mengurus surat ijin usaha dan sebagainya, telah/sedang dilakukan.

- c. Mereka yang tak punya pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan.

- d. Mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja.

(Sumber : "An ILO Manual on Concepts and Methods")

#### **2.12.9 KILM 11. Pengangguran dan Pendidikan**

Melalui indikator ini bisa diperoleh informasi pengangguran berdasarkan pendidikan tertinggi yang ditamatkan. **Pendidikan tertinggi yang ditamatkan** adalah tingkat pendidikan yang dicapai seseorang setelah mengikuti pelajaran pada kelas tertinggi suatu tingkatan sekolah dengan mendapatkan tanda tamat (ijazah). Data TPT dibedakan menurut tiga kelompok tingkat pendidikan :

1. Pendidikan Dasar ke Bawah : Mencakup penduduk yang tidak pernah sekolah, tidak tamat SD/ sederajat dan tamat SD/ sederajat.
2. Pendidikan Menengah : Mencakup penduduk yang tamat SMP/ sederajat dan tamat SMU/ sederajat.
3. Pendidikan Tinggi : Mencakup penduduk yang tamat Perguruan Tinggi.

#### **2.12.10 KILM 12. Setengah Penganggur**

Mereka yang dikategorikan dalam setengah penganggur adalah mereka yang jumlah jam kerjanya di bawah ambang batas jam kerja normal (bekerja kurang dari 35 jam dalam seminggu yang lalu), dengan kondisi:

- a. Mereka yang dengan sukarela mencari pekerjaan tambahan untuk menambah jam kerjanya dari pekerjaannya yang sekarang atau mendapat ganti dari pekerjaannya yang sekarang dengan pekerjaan lain yang mempunyai jam kerja lebih banyak.
- b. Mereka yang tidak mencari pekerjaan tambahan, tetapi bersedia menerima pekerjaan tambahan.

*Halaman ini sengaja dikosongkan*

## BAB 3

### METODOLOGI PENELITIAN

Penelitian ini merupakan kajian terapan dari dua penelitian sebelumnya yang dilakukan oleh Damayanti (2015) dan Siagian (2014). Pengelompokan menggunakan MBC dengan kriteria ICL dilakukan dengan bantuan *package teigen software* R, sedangkan pengelompokan MBC dengan kriteria MML menggunakan *software matlab* dengan menerapkan program RMBC-MML yang terdapat pada penelitian Siagian (2014).

#### 3.1 Sumber Data

Data yang digunakan di dalam penelitian ini adalah data sekunder dari BPS yaitu data Indikator Pasar Tenaga Kerja (KILM). Data bersumber dari hasil Survei Angkatan Kerja Nasional (Sakernas) yang dilakukan rutin pada Februari dan Agustus tahun 2012 sampai 2015.

#### 3.2 Variabel Penelitian

Penelitian ini menggunakan 10 indikator KILM yang terdiri 25 variabel indikator seperti dirinci oleh Tabel 3.1

Tabel 3.1 Variabel Penelitian

Indikator Pasar Tenaga Kerja (KILM)		Variabel Indikator (Persen)
1	Tingkat Partisipasi Angkatan Kerja-TPAK (KILM 1)	Tingkat Partisipasi Angkatan Kerja – TPAK ( $X_1$ )
2	Rasio Penduduk yang Bekerja Terhadap Jumlah Penduduk (KILM 2)	Rasio Penduduk yang Bekerja Terhadap Jumlah Penduduk – EPR ( $X_2$ )
3	Penduduk yang Bekerja Menurut Status Pekerjaan Utama (KILM 3)	1. Persentase Bekerja Status Karyawan ( $X_3$ )
		2. Persentase Bekerja Status Pengusaha ( $X_4$ )

Tabel 3.1 Variabel Penelitian (lanjutan)

<b>Indikator Pasar Tenaga Kerja (KILM)</b>		<b>Variabel Indikator (Persen)</b>
		3. Persentase Bekerja Status Berusaha Sendiri atau Dibantu Buruh Tidak Tetap/ Tidak Dibayar ( $X_5$ )
		4. Persentase Bekerja Status Pekerja Bebas ( $X_6$ )
		5. Persentase Bekerja Status Pekerja Keluarga/tak dibayar ( $X_7$ )
		6. Persentase Bekerja Status Pekerja Rentan ( $X_8$ )
4	Penduduk yang Bekerja Menurut Lapangan Usaha (KILM 4)	1. Persentase Bekerja di Sektor Pertanian ( $X_9$ )
		2. Persentase Bekerja di Sektor Manufaktur ( $X_{10}$ )
		3. Persentase Bekerja di Sektor Jasa-Jasa ( $X_{11}$ )
5	Pekerja Paruh Waktu (KILM 5)	Tingkat Pekerja Paruh Waktu ( $X_{12}$ )
6.	Penduduk yang Bekerja Menurut Jam Kerja (KILM 6)	1. Persentase Bekerja 1-14 Jam ( $X_{13}$ )
		2. Persentase Bekerja 15-24 Jam ( $X_{14}$ )
		3. Persentase Bekerja 25-34 Jam ( $X_{15}$ )
		4. Persentase Bekerja 35-39 Jam ( $X_{16}$ )
		5. Persentase Bekerja 40-48 Jam ( $X_{17}$ )
7.	Penduduk yang Bekerja di Sektor Informal (KILM 7)	Persentase Penduduk Bekerja di Sektor Informal ( $X_{18}$ )
8.	Pengangguran (KILM 8)	Tingkat Pengangguran Terbuka-TPT ( $X_{19}$ )
9.	Pengangguran dan Pendidikan (KILM 11)	1. TPT Pendidikan Dasar ke Bawah ( $X_{20}$ )
		2. TPT Pendidikan Menengah ( $X_{21}$ )
		3. TPT Pendidikan Tinggi ( $X_{22}$ )
10	Setengah Penganggur (KILM 12)	1. Persentase Setengah Penganggur Pendidikan Dasar ke Bawah( $X_{23}$ )
		2. Persentase Setengah Penganggur Pendidikan Menengah ( $X_{24}$ )
		3. Persentase Setengah Penganggur Pendidikan Tinggi ( $X_{25}$ )

25 variabel indikator KILM di atas dibagi menjadi beberapa subset data penelitian yang merupakan kombinasi beberapa variabel indikator pasar tenaga kerja (subset data KILM). Subset data KILM yang terbentuk digunakan sebagai dasar pengelompokan 33 provinsi di Indonesia. Pembentukan subset data KILM

berdasarkan karakteristik pengelompokan yang ingin diteliti, disesuaikan dengan kebutuhan analisis penelitian. Kombinasi variabel  $k_1$  memuat 4 KILM dan terdiri dari 4 variabel indikator. Kombinasi variabel  $k_2$  memuat 1 KILM dan terdiri dari 3 variabel indikator. Demikian seterusnya sampai dengan kombinasi variabel  $k_5$  yang memuat 3 KILM dan terdiri dari 3 variabel indikator.

Tabel 3.2 Kombinasi Variabel pada Subset data KILM

Kombinasi (k)	Variabel Indikator	Karakteristik Pengelompokan (Persen)
$k_1$	$X_1, X_{12}, X_{18}, X_{19}$	TPAK, Pekerja paruh waktu, Pekerja sektor informal, dan TPT
$k_2$	$X_9, X_{10}, X_{11}$	Bekerja menurut lapangan usaha
$k_3$	$X_3, X_4, X_5, X_6, X_7, X_{20}, X_{21}, X_{22}$	Bekerja menurut status, dan TPT menurut pendidikan
$k_4$	$X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{23}, X_{24}, X_{25}$	Bekerja menurut jam kerja, dan Pekerja setengah penganggur
$k_5$	$X_2, X_8, X_{18}$	EPR, Pekerja rentan, dan Pekerja sektor informal

Lima kombinasi variabel indikator tersebut akan dibedakan menurut 8 periode Sakernas yaitu Februari 2012 sampai Agustus 2015. Akibatnya, total subset data yang akan digunakan pada penelitian ini sebanyak 40 subset data KILM (**Lampiran 2**). Nama subset data KILM menunjukkan kombinasi variabel indikator dan periode Sakernas seperti sebagai berikut.

$k_1$  0212 : pengelompokan dilakukan berdasarkan kombinasi variabel indikator  $k_1$  menggunakan data Sakernas Februari 2012.

$k_1$  0812 : pengelompokan dilakukan berdasarkan kombinasi variabel indikator  $k_1$  menggunakan data Sakernas Agustus 2012.

$k_5$  0814 : pengelompokan dilakukan berdasarkan kombinasi variabel indikator  $k_5$  menggunakan data Sakernas Agustus 2014.

Demikian seterusnya sampai dengan subset ke-40 :  $k_5$  0815 yaitu pengelompokan berdasarkan kombinasi variabel indikator  $k_5$  menggunakan data Sakernas Agustus 2015.



### 3.3 Tahapan Penelitian

Berdasarkan rumusan dan batasan masalah yang didukung oleh tinjauan kepustakaan yang telah dipaparkan sebelumnya, maka upaya untuk mencapai tujuan penelitian akan dilakukan melalui tahapan berikut:

**Tujuan 1 :** Pemilihan subset data penelitian menurut indikator pasar tenaga kerja Indonesia yang memenuhi asumsi distribusi  $t$  multivariat.

Untuk mencapai tujuan 1 ini dilakukan langkah-langkah sebagai berikut.

1. Persiapan data dengan memilih indikator KILM sebagai variabel penelitian seperti pada tabel 3.1.
2. Mengkombinasikan beberapa variabel penelitian menjadi kombinasi variabel seperti pada tabel 3.2.
3. Kombinasi variabel pada poin (2) diterapkan pada delapan periode pencacahan Sakernas sehingga terbentuk 40 subset data KILM (**Lampiran 2**).
3. Masing-masing subset data KILM dilakukan pemeriksaan keberadaan *outlier* menggunakan ukuran jarak *robust* (*robust distance*).
4. Masing-masing subset data KILM dilakukan pengujian distribusi  $t$  multivariat dengan menghitung ukuran kelandaian (*kurtosis*) dan kemencengan (*skewness*) data.
5. Subset subset data KILM yang memenuhi asumsi distribusi  $t$  multivariat akan digunakan sebagai dasar pengelompokan (*clustering*) provinsi di Indonesia.

**Tujuan 2 :** Membentuk kelompok provinsi di Indonesia menurut indikator pasar tenaga kerja menggunakan *model-based clustering t* multivariat dengan kriteria ICL dan MML.

Tujuan 2 dilakukan pada setiap subset data KILM yang memenuhi tujuan 1 menurut langkah-langkah sebagai berikut.

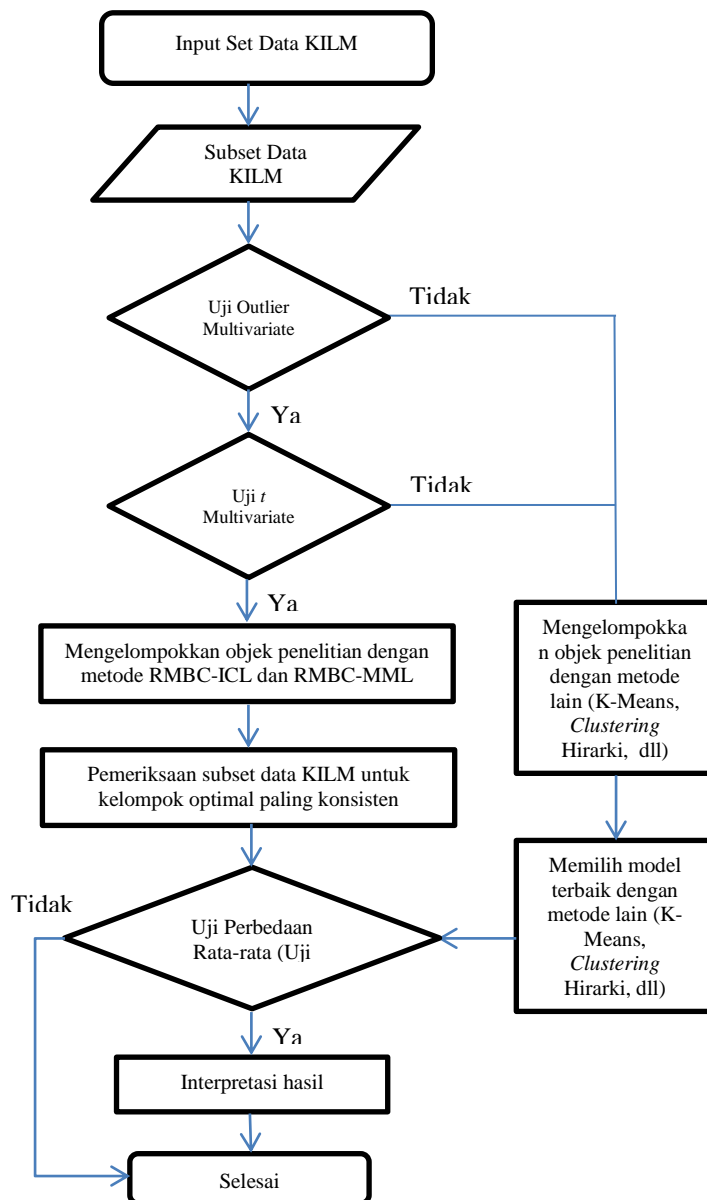
1. Pembentukan kelompok optimal dengan kriteria ICL. Proses pengelompokan menggunakan *software R package*.
2. Pembentukan kelompok optimal dengan kriteria MML. Proses pengelompokan menggunakan *software matlab*.
3. Berdasarkan hasil pengelompokan dengan kriteria ICL dan MML,

dilakukan pemeriksaan subset data KILM yang menghasilkan kelompok optimal paling konsisten.

4. Melakukan uji perbedaan rata-rata terhadap kelompok yang terbentuk dari subset data terpilih pada poin (3) menggunakan uji Manova.

**Tujuan 3** : Analisis pasar tenaga kerja Indonesia.

Secara singkat, tahapan penelitian ditunjukkan oleh diagram alir berikut ini.



Gambar 3.1 Diagram Alir Tahapan Penelitian

*Halaman ini sengaja dikosongkan*

## **BAB 4**

### **HASIL DAN PEMBAHASAN**

Seperti disampaikan pada bab 1, tujuan utama penelitian ini adalah mengetahui provinsi-provinsi di Indonesia yang memiliki kemiripan karakteristik berdasarkan hasil pengelompokan (*clustering*) menurut beberapa indikator pasar tenaga kerja (multivariat). Metode pengelompokan adalah *model-based clustering* dengan kriteria ICL dan MML. Hasil dan pembahasan penelitian diperoleh dengan melakukan urutan tahapan penelitian sebagaimana telah disampaikan pada bab 3 sebelumnya.

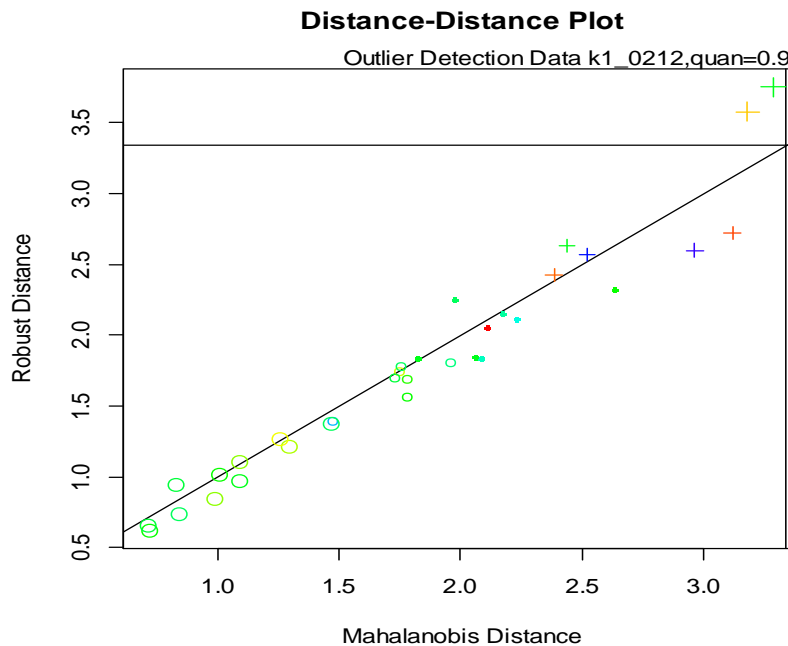
#### **4.1 Subset Data Penelitian Menurut Indikator Pasar Tenaga Kerja Indonesia yang Memenuhi Asumsi Distribusi $t$ Multivariat**

Data yang digunakan pada penelitian ini harus memenuhi asumsi distribusi  $t$  multivariat, sehingga metode MBC dengan distribusi  $t$  multivariat bisa diterapkan. Pemeriksaan subset data yang digunakan untuk mengelompokkan provinsi di Indonesia dilakukan melalui dua tahapan :

1. Deteksi *outlier* multivariat
2. Pengujian asumsi distribusi

##### **4.1.1 Deteksi *Outlier* Multivariat**

Deteksi *outlier* multivariat pada data penelitian dilakukan untuk membuktikan asumsi awal bahwa data indikator pasar tenaga kerja di Indonesia cenderung memuat *outlier*, sehingga *model-based clustering finite mixture  $t$*  multivariat tepat diterapkan untuk mendapatkan hasil *clustering* yang *robust*. Proses deteksi *outlier* multivariat pada data penelitian diawali dengan menghitung jarak mahalanobis dan jarak *robust* untuk semua objek pengamatan. Kemudian membandingkannya dengan kuantil dari distribusi  $\chi_{p;0.975}$  sebagai nilai *cut off*. Nilai *cut off* subset data KILM bisa berbeda-beda, tergantung jumlah variabel ( $p$ ) masing-masing subset data.



Gambar 4.1 *Plot* Jarak Mahalanobis Terhadap Jarak *Robust* Subset Data  $k_1$  0212

Gambar 4.1 menyajikan *plot* jarak mahalanobis terhadap jarak *robust* subset data  $k_1$  0212 melalui fungsi *distance-distance plot* (*dd plot*) dari program R *package mvoutlier*. Jika data tidak terkontaminasi *outlier* maka seluruh titik akan terletak di sekitar garis lurus, perpotongan jarak mahalanobis dan jarak *robust*. Selain berfungsi sebagai sebuah diagnostik *plot* deteksi *outlier* multivariat, *dd plot* juga dapat digunakan untuk mendiagnosa asumsi distribusi normal multivariat dan *elliptical symetri* (Rousseeuw & Van Driessen, 1999). *Plot* di atas menunjukkan bahwa terdapat dua provinsi yang terdeteksi sebagai *outlier* pada Sakernas february 2012. Kedua provinsi tersebut adalah DKI Jakarta dan Bali, sebagaimana ditunjukkan *output dd-plot* terlampir (Lampiran 12). Selain deteksi *outlier* melalui *plot*, *dd-plot* juga menyajikan *output* nilai jarak *robust* dan mahalanobis 33 provinsi. Keterangan “TRUE” berarti provinsi terdeteksi sebagai *outlier* pada subset data KILM  $k_1$  0212. Sebaliknya, keterangan “FALSE” menunjukkan bahwa provinsi tidak terdeteksi sebagai *outlier*.

Tabel 4.1 Jumlah Provinsi Terdeteksi Sebagai Outlier Pada Kuantil 90 Persen

Kombinasi	<u>Sakernas Februari</u>				<u>Sakernas Agustus</u>			
	2012	2013	2014	2015	2012	2013	2014	2015
$k_1$	2	1	3	3	3	1	3	2
$k_2$	2	2	3	1	3	3	1	1
$k_3$	3	3	3	3	3	3	3	3
$k_4$	3	3	3	3	3	3	3	3
$k_5$	3	3	3	3	3	3	3	2

Dari Tabel 4.1 diketahui semua subset data KILM memuat beberapa provinsi yang teridentifikasi sebagai *outlier* pada kuantil 90 persen. Berdasarkan hasil pengolahan *dd-plot* dengan *package mvoutlier* (Lampiran 13 sampai dengan Lampiran 17). Papua merupakan provinsi yang sering terdeteksi sebagai *outlier* jika diukur berdasarkan kombinasi variabel  $k_1$  (TPAK, Pekerja paruh waktu, Pekerja sektor informal, dan TPT). DKI Jakarta menjadi provinsi yang selalu terdeteksi sebagai *outlier* dari Sakernas february 2012 sampai dengan Sakernas agustus 2015 menurut kombinasi variabel  $k_2$  (Bekerja menurut lapangan usaha). Jika diukur menurut kombinasi variabel  $k_3$  (Bekerja menurut status, dan TPT menurut pendidikan), provinsi yang terdeteksi sebagai *outlier* cenderung bervariasi atau berbeda-beda pada setiap pencacahan Sakernas. DKI Jakarta, Sulawesi Barat dan Papua merupakan provinsi yang dominan terdeteksi sebagai *outlier* jika dilihat berdasarkan kombinasi variabel  $k_4$  (Bekerja menurut jam kerja, dan Pekerja setengah penganggur). Jika diukur menurut kombinasi variabel  $k_5$  (EPR, Pekerja rentan, dan Pekerja sektor informal), Kepulauan Riau, DKI Jakarta, Bali dan Papua merupakan provinsi yang beberapa kali terdeteksi sebagai *outlier*. Provinsi-provinsi yang terdeteksi sebagai *outlier* perlu menjadi perhatian khusus saat pengecekan hasil pengelompokan.

#### 4.1.2 Pengujian Asumsi Distribusi

Pengujian asumsi dilakukan untuk memastikan apakah asumsi awal mengenai distribusi data penelitian bisa terpenuhi atau tidak. Di latar belakang telah disampaikan bahwa data indikator pasar tenaga kerja Indonesia diasumsikan cenderung memuat *outlier* dan tidak berdistribusi normal. Menggunakan *software*

R package *psych*, dilakukan *mardia test* untuk uji asumsi normal multivariat dengan hipotesis :

$H_0$  : Data berdistribusi normal multivariat

$H_1$  : Data tidak berdistribusi normal multivariat.

Keputusan tolak  $H_0$  jika  $p - value < \alpha = 0.05$ . Selain ditentukan oleh nilai  $p - value$ , keputusan tolak  $H_0$  juga bisa berdasarkan nilai kemencengan/*skewness* ( $b_{1p}$ )  $> 0$  dan nilai kelandaian/*kurtosis* ( $b_{2p}$ )  $> p(p + 2)$ . Dimana  $p$  menunjukkan banyak variabel. Oleh karena penelitian ini menggunakan *model-based clustering* dengan distribusi *t multivariate*, maka diharapkan keputusan yang diperoleh adalah **tolak  $H_0$**  dan mengindikasikan data penelitian **tidak berdistribusi normal**.

Berdasarkan output *mardia test* (**Lampiran 18**), diketahui kemencengan subset data  $k_1$  pada Februari 2012 memiliki nilai *probability* ( $p$ -value) = 0.087 (lebih besar dari  $\alpha = 0.05$ ). Namun, untuk *small sample skewness* memiliki nilai *probability* = 0.035 (lebih kecil dari  $\alpha$ ) sehingga diperoleh keputusan **tolak  $H_0$** . Ini berarti subset data  $k_1$  0212 tidak berdistribusi normal multivariat. Hasil uji *mardia* subset data  $k_1$  pada Februari 2013, 2014 dan 2015 memiliki nilai *probability skewness* dan *small sample skewness* yang lebih besar dari  $\alpha$ , sehingga keputusan : **gagal tolak  $H_0$** . Ini berarti subset data  $k_1$  pada tiga periode Sakernas ini berdistribusi normal multivariat. Dengan cara yang sama dilakukan uji *mardia* terhadap subset subset data yang lain, dan diperoleh hasil uji seperti pada **Lampiran 19**. Dari 40 subset data KILM diketahui kombinasi variabel  $k_1$  pada beberapa periode pencacahan Sakernas berdistribusi normal multivariat, sedangkan kombinasi variabel lain ( $k_2$ ,  $k_3$ ,  $k_4$ , dan  $k_5$ ) cenderung menceng. Oleh karena itu, pengelompokan dengan *model-based clustering t multivariate* bisa diaplikasikan pada kombinasi variabel  $k_2$ ,  $k_3$ ,  $k_4$ , dan  $k_5$ .

## 4.2 Pengelompokan Provinsi di Indonesia Menurut Indikator Pasar Tenaga Kerja

Berdasarkan hasil deteksi *outlier* dan uji asumsi distribusi subset data KILM, pengelompokan provinsi di Indonesia akan dilakukan menurut empat kombinasi variabel indikator pasar tenaga kerja sebagai berikut :

1. Pengelompokan berdasarkan kombinasi variabel indikator **Bekerja Menurut Lapangan Usaha ( $k_2$ )**.
2. Pengelompokan berdasarkan kombinasi variabel indikator **Bekerja Menurut Status, dan TPT Menurut Pendidikan ( $k_3$ )**.
3. Pengelompokan berdasarkan kombinasi variabel indikator **Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur ( $k_4$ )**.
4. Pengelompokan berdasarkan kombinasi variabel indikator **EPR, Pekerja Rentan, dan Pekerja Sektor Informal ( $k_5$ )**.

Melalui pengelompokan menggunakan *model based-clustering* dengan kriteria ICL dan MML, diharapkan bisa diperoleh kelompok optimal yang menggambarkan dengan baik kondisi pasar tenaga kerja Indonesia.

### 4.2.1 Pengelompokan Provinsi di Indonesia menggunakan *Model- Based Clustering* dengan Kriteria *Integrated Completed Likelihood* (MBC-ICL)

*Package teigen software* R mampu mengidentifikasi 28 model yang mungkin dengan jumlah kelompok maksimal 9 kelompok untuk MBC *mixture t* multivariat dengan kriteria ICL.

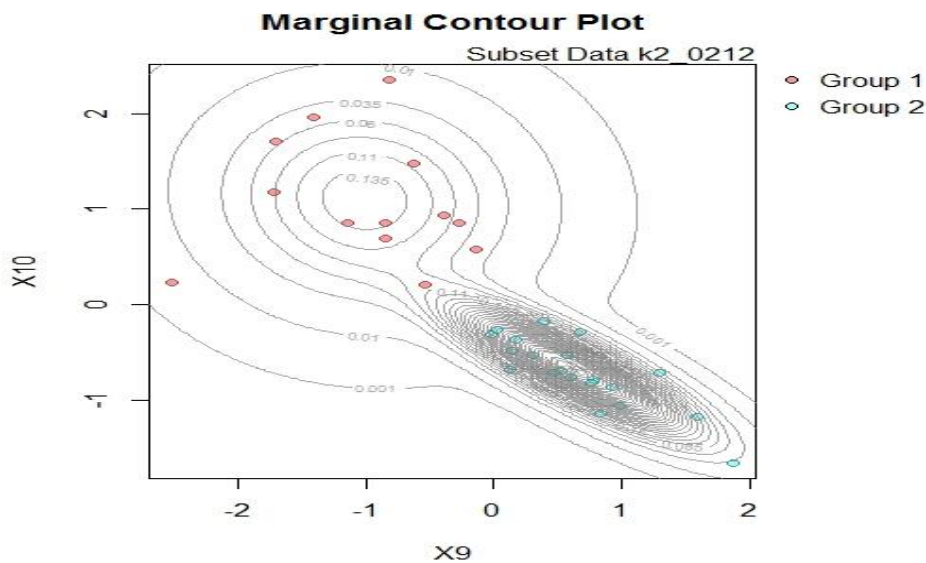
Lampiran 1 merinci tata nama model hasil pengelompokan MBC menggunakan kriteria ICL. Pengelompokan provinsi berdasarkan subset data  $k_2$  0212 mungkin dilakukan sampai dengan 4 kelompok (**Lampiran 21**). Lebih dari itu, proses iterasi menunjukkan hasil **infinite** (tak terbatas). Pemilihan model terbaik dan jumlah kelompok optimal ditentukan melalui **nilai ICL terbesar**.

Output pada **Lampiran 21** menunjukkan bahwa model terbaik pada pengelompokan provinsi di Indonesia menggunakan subset data  $k_2$  0212 adalah UUUC dengan jumlah kelompok optimal sama dengan 2. *Software R* dengan



*package teigen* melalui command “`teigen(data)$iclresult`” tidak hanya menampilkan kemungkinan model terbaik dan jumlah kelompok optimal, tetapi juga hasil klasifikasi untuk anggota kelompok (**Lampiran 20**) dan estimasi parameter kelompok optimal (**Lampiran 22**). Kelompok 1 terdiri dari 13 provinsi, sedangkan kelompok 2 beranggotakan 20 provinsi.

Berdasarkan persamaan (2.5), model terbaik UUUC menunjukkan bahwa model matriks kovarians kedua kelompok adalah  $\Sigma_g = \lambda_g D_g A_g D_g^T$ . Ini berarti kedua kelompok memiliki volume elips, orientasi, dan kontur fungsi kepadatan yang berbeda tetapi derajat bebas kedua kelompok bernilai sama. Perbedaan volume elips, orientasi dan kontur menyebabkan kedua kelompok memiliki matriks varians kovarians ( $\Sigma_g$ ) yang berbeda (**Lampiran 22**). Akibatnya, pengelompokan Provinsi di Indonesia berdasarkan kombinasi variabel indikator bekerja menurut lapangan usaha ( $k_2$ ) data Sakernas Februari 2012 memiliki *marginal contour plot* seperti Gambar 4.2 berikut.



Gambar 4.2 *Marginal Contour Plot* Subset Data  $k_2$  0212

Dengan cara yang sama bisa diperoleh pengelompokan untuk subset-subset data yang lain. Jumlah kelompok optimal yang dihasilkan pada tiap subset data penelitian dirangkum pada **Tabel 4.2** berikut.

Tabel 4.2 Jumlah Kelompok Optimal MBC-ICL

Kombinasi	<u>Sakernas Februari</u>				<u>Sakernas Agustus</u>			
Variabel	2012	2013	2014	2015	2012	2013	2014	2015
k <sub>2</sub>	2	2	2	2	1	1	1	2
k <sub>3</sub>	1	2	1	1	2	1	2	1
k <sub>4</sub>	1	1	4	2	1	1	1	1
k <sub>5</sub>	1	2	2	1	1	1	1	1

Sebagian besar pengelompokan pada Sakernas Agustus menghasilkan kelompok optimal satu kelompok. Ini berarti karakteristik pasar tenaga kerja 33 provinsi di Indonesia terlihat sama pada periode Sakernas tersebut. Kondisi ini perlu dikaji lebih lanjut, agar tidak terjadi kesalahan interpretasi. Apalagi jika sebaran data antar kedua periode Sakernas tidak jauh berbeda. Salah satu hasil pengelompokan menggunakan jumlah kelompok optimal dengan kriteria ICL dapat dilihat pada **Tabel 4.3**.

Tabel 4.3 Anggota Kelompok Provinsi di Indonesia Berdasarkan Subset Data k<sub>2</sub> 0212 Menggunakan MBC-ICL

Kelompok 1	Kelompok 2
Kep. Bangka Belitung	Aceh
Kep. Riau	Sumatera utara
DKI Jakarta	Sumatera barat
Jawa Barat	Riau
Jawa Tengah	Jambi
DI Yogyakarta	Sumatera Selatan
Jawa Timur	Bengkulu
Banten	Lampung
Bali	Nusa Tenggara Barat
Kalimantan Selatan	Nusa Tenggara Timur
Kalimantan Timur	Kalimantan Barat
Sulawesi Utara	Kalimantan Tengah
Gorontalo	Sulawesi Tengah
	Sulawesi Selatan
	Sulawesi Tenggara
	Sulawesi Barat
	Maluku
	Maluku utara
	Papua barat
	Papua

Berdasarkan statistik deskriptif subset data k<sub>2</sub> 0212 pada **Lampiran 24** diketahui bahwa sebagian besar penduduk di provinsi pada kelompok 1 bekerja di sektor jasa-jasa dan manufaktur. Sedangkan kelompok 2 penduduknya lebih banyak bekerja di sektor pertanian.

#### 4.2.2 Pengelompokan Provinsi di Indonesia Menggunakan *Robust Model Based Clustering* dengan Kriteria *Minimum Message Length* (RMBC-MML)

Pengelompokan menggunakan kriteria MML diperoleh dengan menerapkan algoritma RMBC-MML seperti pada penelitian Siagian (2014). Berikut ditulis ulang **algoritma metode RMBC-MML** sebagai alat kalkulasi numerik untuk proses *clustering* dan penaksiran parameter pada model *finite mixture t* multivariat.

##### 1. Input data set.

Input data set berkaitan dengan output yang akan dihasilkan yaitu jumlah parameter. Jumlah parameter yang harus diestimasi, berhubungan langsung dengan besarnya dimensi ruang data pengamatan. Banyaknya parameter bebas yang harus diestimasi pada model *mixture t* multivariat adalah (Bouveyron & Brunet Saumard, 2014) :

$$n_p = (G - 1) + Gp + \frac{1}{2}Gp(p + 1) + (G \times 1)$$

Keterangan :

$n_p$  : Banyak parameter bebas yang harus diestimasi pada model

$(G - 1)$  : Banyaknya parameter *mixing proportion*

$Gp$  : Banyak parameter *mean*

$\frac{1}{2}Gp(p + 1)$  : Banyak parameter matriks varians kovarians

$(G \times 1)$  : Banyak parameter derajat bebas

##### 2. Tahap inisialisasi.

2.1. Aplikasikan metode *clustering* sederhana yaitu *K-means* sebagai inisialisasi nilai parameter untuk *mean* ( $\mu$ ) kelompok.

$\mu_1, \mu_2, \dots, \mu_{g\_max}$ , untuk  $g\_max > g\_true\ group$ .

## 2.2. Input inialisasi nilai parameter lain.

Input data set untuk  $P_j$ , Sigma,  $\nu$ , dan  $N$ .  $P_j$  merupakan inisial untuk *mixing proportion*. Sigma yaitu matriks Identitas yang merupakan inisial untuk matriks varians kovarians.  $\nu$  adalah inisial untuk derajat bebas. Pada tesis ini nilainya dibatasi :  $1 \leq \nu \leq N$ . Nilai  $N$  adalah inisial banyaknya parameter dalam 1 kelompok ( $G = 1$ ), bisa dihitung berdasarkan persamaan  $n_p$  di atas. Dengan asumsi antara variabel saling bebas, matriks varians kovarians sama dengan matriks varians, maka diperoleh nilai  $N$  :

$$N = (1 - 1) + (1)p + (1)p + (1 \times 1) = p + p + 1 = 2p + 1$$

## 3. Tahap reduksi dimensi dengan ROBPCA

### 4. Algoritma EM

Lakukan secara berulang langkah E dan langkah M sampai konvergen ( $\Delta \ln L_{c,p}(\vartheta) < 0,1$  dan  $\Delta \nu_g^{k+1} < 10^{-2}$ ) untuk  $G > g_{\min}$

#### 4.1. Langkah E (*Expectation*)

Update nilai probabilitas posterior  $\hat{t}_{ig}^k$  dan bobot  $\hat{u}_{ig}^k$  setiap obyek pengamatan untuk setiap kelompok, yaitu :

$$\hat{t}_{ig}^k = \frac{\hat{\pi}_g^k f_g(x_i; \mu_g^k, \Sigma_g^k, \nu_g^k)}{\sum_{g=1}^G \hat{\pi}_g^k f_g(x_i; \mu_g^k, \Sigma_g^k, \nu_g^k)}$$

$$\hat{u}_{ig}^k = \frac{\hat{\nu}_g^k + p}{\hat{\nu}_g^k + \delta(x_i, \hat{\mu}_g^k; \hat{\Sigma}_g^k)}$$

#### 4.2. Langkah M (*Maximization*)

a. Update *mixing proportion*  $\hat{\pi}_g^{(k+1)}$  setiap kelompok sampai  $\left| \sum_{g=1}^G \hat{\pi}_g^{(k+1)} - 1 \right| < 10^{-4}$  dan keluarkan purge kelompok dengan nilai *mixing proportion*  $\hat{\pi}_g^{k+1} = 0$ .

$$\hat{\pi}_g^{k+1} = \frac{1}{\left(n - \frac{G \cdot m_g}{2}\right)} \left( \sum_{i=1}^n \hat{t}_{ig}^{(k)} - \frac{m_g}{2} \right)$$

Ukuran  $G$  adalah  $\#(cardinality)$  dari  $\hat{\pi}_g^{(k+1)} > 0$ .

- b. *Update mean* setiap kelompok
- c. *Update matriks varians kovarians* setiap kelompok
- d. *Update derajat bebas*
- e. *Update nilai*  $f_g(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_g^k, \hat{\boldsymbol{\Sigma}}_g^k, \hat{v}_g^k)$
- f. *Update nilai*  $\ln L_{c,p}(\vartheta)$

Jika  $\ln L_{c,p_{max}}(\vartheta) = \ln L_{c,p}(\vartheta)$  maka putuskan parameter

$$\{\hat{\pi}_g^{(k+1)}, \hat{\boldsymbol{\mu}}_g^{(k+1)}, \hat{\boldsymbol{\Sigma}}_g^{(k+1)}, \hat{v}_g^{(k+1)}\}$$

sebagai parameter optimal.

Penulisan program RMBC-MML dibuat dengan MATLAB dan secara rinci sintaksnya disajikan pada Lampiran 29.

Tabel 4.4 Jumlah Kelompok Optimal RMBC-MML

Kombinasi	<u>Sakernas Februari</u>				<u>Sakernas Agustus</u>			
Variabel	2012	2013	2014	2015	2012	2013	2014	2015
k <sub>2</sub>	2	2	2	1	2	2	1	1
k <sub>3</sub>	*	*	*	*	1	1	*	*
k <sub>4</sub>	*	*	*	*	1	*	*	1
k <sub>5</sub>	1	2	1	1	1	1	2	2

Keterangan : tanda \* artinya pada subset tersebut tidak bisa terbentuk kelompok atau terjadi *error* saat diolah menggunakan program RMBC-MML terlampir.

Pada **Tabel 4.4** diketahui subset data dengan jumlah variabel 8 cenderung tidak berhasil dilakukan pengelompokan (subset data k<sub>3</sub> dan k<sub>4</sub>). Subset data k<sub>2</sub> dan k<sub>5</sub> dengan jumlah variabel yang lebih sedikit berhasil dilakukan pengelompokan. Penelitian dengan data berdimensi besar memiliki masalah *curse of dimensionality*, suatu istilah yang diperkenalkan oleh R. Bellman untuk menggambarkan betapa sulitnya bekerja pada ruang data berdimensi besar (Bouveyron & Brunet Saumard, 2014). Penelitian Siagian (2014) berhasil mengaplikasikan RMBC-MML dengan 10 variabel indikator kerawanan sosial, pada sampel besar yaitu mengelompokkan 497 kabupaten/kota di Indonesia. Pada saat dilakukan simulasi pada variabel k<sub>3</sub> dan k<sub>4</sub> dengan menambahkan jumlah

sampel ( $n=33$  menjadi  $n=132$ ), tidak terjadi *error* dan berhasil terbentuk 2 kelompok atau lebih. Hal ini mengindikasikan bahwa *error* terjadi dikarenakan jumlah sampel yang digunakan pada subset data  $k_3$  dan  $k_4$  kecil, sedangkan variabelnya banyak.

Tabel 4.5 Anggota Kelompok Provinsi di Indonesia Berdasarkan Subset Data  $k_2$  0212 Menggunakan MBC-MML

Kelompok 1	Kelompok 2
Kep. Bangka Belitung	Aceh
Kep. Riau	Sumatera Utara
DKI Jakarta	Sumatera Barat
Jawa Barat	Riau
Jawa Tengah	Jambi
DI Yogyakarta	Sumatera Selatan
Jawa Timur	Bengkulu
Banten	Lampung
Bali	Nusa Tenggara Barat
Kalimantan Selatan	Nusa Tenggara Timur
Kalimantan Timur	Kalimantan Barat
Sulawesi Utara	Kalimantan Tengah
Gorontalo	Sulawesi Tengah
	Sulawesi Selatan
	Sulawesi Tenggara
	Sulawesi Barat
	Maluku
	Maluku Utara
	Papua Barat
	Papua

Pengelompokan provinsi di Indonesia menggunakan MBC dengan kriteria MML pada subset data  $k_2$  0212 menghasilkan anggota kelompok yang sama dengan kriteria ICL. Hasil pengelompokan dengan kriteria MML juga menggambarkan bahwa tenaga kerja yang berada di Provinsi Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta dan beberapa provinsi lain kelompok 1 lebih banyak yang bekerja di sektor jasa-jasa dan manufaktur dibandingkan dengan provinsi-provinsi yang berada di kelompok 2 seperti Aceh, Lampung, Papua dan sebagainya. Provinsi di kelompok 2 lebih banyak bekerja di sektor Pertanian dibandingkan bekerja di sektor jasa-jasa dan manufaktur.

### 4.3 Kelompok Optimal MBC-ICL dan RMBC-MML

Kelebihan pada *model-based clustering* yang tidak dimiliki metode *clustering* klasik lainnya yaitu referensi jumlah kelompok optimal. Kelompok optimal dipilih dari model *cluster* yang memiliki nilai kriteria terbesar. Meskipun demikian, saat melakukan analisis juga perlu dikaji ukuran dispersi (penyebaran) kelompok optimal seperti nilai koefisien variasi dan *range* kelompok (nilai maksimum dan nilai minimum).

Kedua kriteria ICL dan MML menghasilkan jumlah kelompok optimal yang sama ( $G=2$ ) pada subset data  $k_2$  0212,  $k_2$  0213,  $k_2$  0214, dan  $k_5$  0213.

Tabel 4.6 Klasifikasi Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Lapangan Usaha ( $k_2$  0214) Menggunakan MBC-ICL dan MBC-MML (Kelompok Optimal)

Kelompok 1		Kelompok 2	
1	Aceh	1	Kep. Bangka Belitung
2	Sumatera Utara	2	Kep. Riau
3	Sumatera Barat	3	Jawa Barat
4	Riau	4	Jawa Tengah
5	Jambi	5	Jawa Timur
6	Sumatera Selatan	6	Banten
7	Bengkulu	7	Bali
8	Lampung	8	Kalimantan Selatan
9	DKI Jakarta		
10	DI Yogyakarta		
11	Nusa Tenggara Barat		
12	Nusa Tenggara Timur		
13	Kalimantan Barat		
14	Kalimantan Tengah		
15	Kalimantan Timur		
16	Sulawesi Utara		
17	Sulawesi Tengah		
18	Sulawesi Selatan		
19	Sulawesi Tenggara		
20	Gorontalo		
21	Sulawesi Barat		
22	Maluku		
23	Maluku Utara		
24	Papua Barat		
25	Papua		

Selain jumlah kelompok optimal yang sama, keanggotaan kelompok juga sama untuk subset data  $k_2$  0212 dan  $k_2$  0214. Sedangkan pada subset data  $k_2$  0213

terdapat perbedaan pengelompokan untuk Provinsi DKI Jakarta, DI Yogyakarta dan Kalimantan Timur (**Lampiran 33**). Pengelompokan dengan kriteria ICL menunjukkan terjadi pergeseran struktur keanggotaan ketiga provinsi ini dibandingkan dengan tahun 2012. Sedangkan kriteria MML tetap mengelompokkan di kelompok yang sama dengan tahun 2012. Akan tetapi, ketiga provinsi tersebut kemudian dikelompokkan ke kelompok 1 oleh kriteria MML pada tahun 2014.

Berdasarkan **Tabel 4.6** diketahui bahwa Provinsi DKI Jakarta, DI Yogyakarta, dan Kalimantan Timur berada pada kelompok provinsi dengan persentase bekerja terbesar di sektor pertanian baik oleh kriteria ICL maupun MML. Padahal data menunjukkan bahwa persentase bekerja di sektor jasa-jasa pada ketiga provinsi ini tahun 2012 sampai 2015 masih tinggi yaitu di atas 50 persen, dan persentase bekerja di sektor pertanian juga masih rendah, tidak mengalami peningkatan yang berarti.

Tabel 4.7 Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Lapangan Usaha (k<sub>2</sub> 0214) Menggunakan MBC-ICL dan MBC-MML (Kelompok Optimal)

Variable	k <sub>2</sub> 0214	Total Count	Mean	SE Mean	Coef Var	Min	Max
X <sub>9</sub>	1	25	45,34	2,87	31,69	2,17	73,43
	2	8	25,78	3,23	35,47	13,96	36,86
X <sub>10</sub>	1	25	12,97	0,78	29,98	3,84	21,38
	2	8	27,19	1,33	13,81	21,48	31,38
X <sub>11</sub>	1	25	41,68	2,22	26,64	22,72	78,35
	2	8	47,03	2,47	14,87	40,44	57,27

Statistik deskriptif **Tabel 4.7** menunjukkan dispersi (penyebaran) antar anggota kelompok pada kelompok 1 masih cukup tinggi dilihat dari nilai koefisien variasi dan *range* kelompok. Ini berarti anggota kelompok 1 masih heterogen (berbeda). *Range* kelompok 1 pada variabel persentase bekerja di sektor pertanian (X<sub>9</sub>) terlalu jauh, dimana nilai minimum 2,17 persen dan maksimum 73,43 persen. Kondisi yang sama terjadi di variabel X<sub>10</sub> dan X<sub>11</sub>. Sedangkan anggota kelompok 2 terlihat lebih homogen. Koefisien variasi kelompok 2 lebih rendah dibandingkan dengan kelompok 1. *Range* kelompok juga tidak selebar kelompok 1. Nilai minimum dan nilai maksimum kelompok 2 pada X<sub>9</sub>, X<sub>10</sub>, dan X<sub>11</sub> tidak berbeda



jauh. Pengelompokan objek menggunakan banyak variabel (multivariat) memerlukan informasi tambahan untuk memastikan bahwa hasil pengelompokan tidak bias.

Hasil pengelompokan MBC-ICL dan RMBC-MML memungkinkan untuk dilakukan analisis pasar tenaga kerja Indonesia kondisi Agustus 2015 berdasarkan:

1. Indikator Bekerja Menurut Lapangan Usaha ( $k_2$ ) menggunakan kriteria MBC-ICL dengan jumlah kelompok optimal  $G = 2$ .
2. Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan ( $k_3$ ) menggunakan kriteria MBC-ICL dengan jumlah kelompok *fixed*  $G=2$ .
3. Indikator Bekerja menurut jam kerja, dan Pekerja setengah penganggur ( $k_4$ ) menggunakan kriteria MBC-ICL dengan jumlah kelompok *fixed*  $G = 2$ .
4. Indikator EPR, Pekerja rentan, dan Pekerja sektor informal ( $k_5$ ) menggunakan kriteria RMBC-MML dengan jumlah kelompok optimal  $G = 2$ .

#### 4.4 Uji Kesamaan Kelompok

Untuk mengetahui apakah kelompok yang dihasilkan memang berbeda signifikan, perlu dilakukan uji beda rata-rata melalui uji Manova. Manova merupakan salah satu analisis multivariat yang digunakan untuk mengetahui apakah vektor rata-rata populasi sama. Jika tidak sama, komponen rata-rata mana yang berbeda secara nyata. Berikut adalah hipotesis untuk uji beda rata-rata kedua kelompok.

$H_0 : \tau_1 = \tau_2 = 0$  (tidak ada perbedaan antar kelompok)

$H_1 : \tau_1 \neq \tau_2$  (kelompok 1 berbeda dengan kelompok 2)

Seperti yang dijelaskan pada subbab 2.8, keputusan Tolak  $H_0$  jika nilai *p-value* lebih kecil dari  $\alpha = 0.05$ , yang berarti kelompok 1 memang berbeda dengan kelompok 2. Pada penelitian ini, pengelompokan provinsi di Indonesia akan dilakukan berdasarkan subset data  $k_2$ ,  $k_3$ ,  $k_4$ , dan  $k_5$  pada data Sakernas Agustus 2015.

Melalui bantuan *software* Minitab diperoleh hasil uji Manova untuk masing-masing subset data.

Tabel 4.8 Hasil Uji Kesamaan Kelompok (Uji Manova) Subset Data  $k_2$  0815,  $k_3$  0815,  $k_4$  0815, dan  $k_5$  0815 dengan Statistik Uji *Pillai's Trace*

Subset	Nilai Pillai's Trace	F	<i>p-value</i>
$K_2$ 0815	0,96146	8,949	0,000
$K_3$ 0815	0,79210	11,430	0,000
$K_4$ 0815	0,73870	8,481	0,000
$K_5$ 0815	0,36344	5,519	0,004

Statistik Pillai's Trace memiliki *p-value* lebih kecil dari 0,05 untuk semua subset data. Ini berarti  $H_0$  ditolak yang menunjukkan adanya perbedaan vektor *mean* antar kelompok di masing-masing subset data sehingga analisis *cluster* untuk provinsi-provinsi di Indonesia berdasarkan kombinasi variabel  $k_2$ ,  $k_3$ ,  $k_4$ , dan  $k_5$  bisa dilakukan.

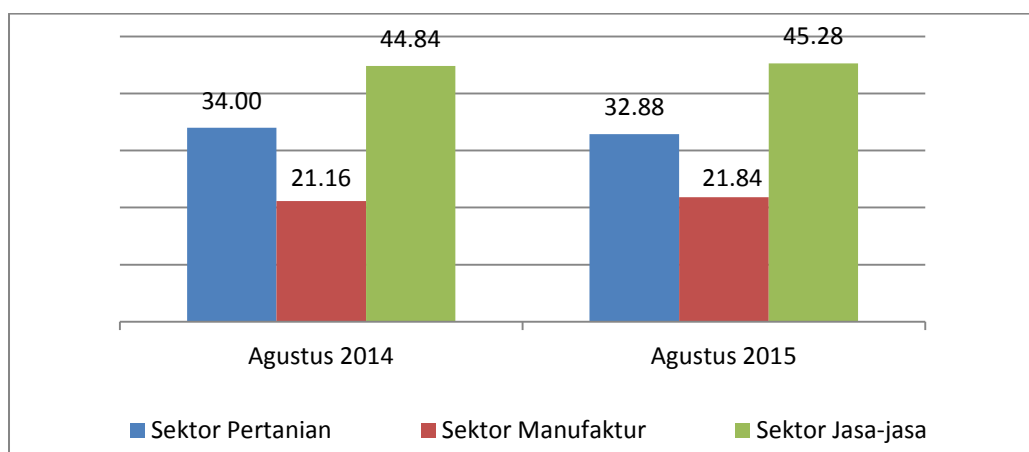
#### 4.5 Analisis Pasar Tenaga Kerja Indonesia

Deskripsi pasar tenaga kerja Indonesia digambarkan dengan cukup lengkap pada publikasi rutin BPS : “Indikator Pasar Tenaga Kerja”. Pada bulan Agustus 2015, jumlah angkatan kerja dan TPAK cenderung menurun. TPAK dan EPR laki-laki masih lebih tinggi dari pada TPAK dan EPR perempuan. Papua dan Bali juga masih menjadi dua provinsi teratas untuk angka TPAK di Indonesia. Jika mayoritas penduduk Bali bekerja di sektor jasa-jasa, maka penduduk Papua mayoritas bekerja di sektor pertanian. Pekerja di Indonesia sebagian besar bekerja di atas 35 jam per minggu. Sampai dengan Agustus 2015, peran sektor informal menggerakkan pasar tenaga kerja lebih besar dibandingkan sektor formal yaitu 51,72 persen, dan terpusat di perdesaan. Papua kembali menjadi provinsi dengan persentase tertinggi penduduk bekerja di sektor informal. Provinsi Aceh, Maluku dan Banten menjadi provinsi dengan TPT tertinggi dan mayoritas berada pada TPT dengan pendidikan Sekolah Menengah. Sedangkan Kalimantan Barat, Jawa Barat dan Banten adalah tiga provinsi dengan persentase pengangguran berpendidikan rendah paling tinggi.

Selanjutnya akan dilakukan analisis *cluster* provinsi-provinsi di Indonesia untuk melihat provinsi-provinsi mana saja yang memiliki kesamaan karakteristik, apakah berada pada kelompok rata-rata tinggi atau kelompok rata-rata rendah.

#### 4.5.1 Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Lapangan Usaha ( $k_2$ )

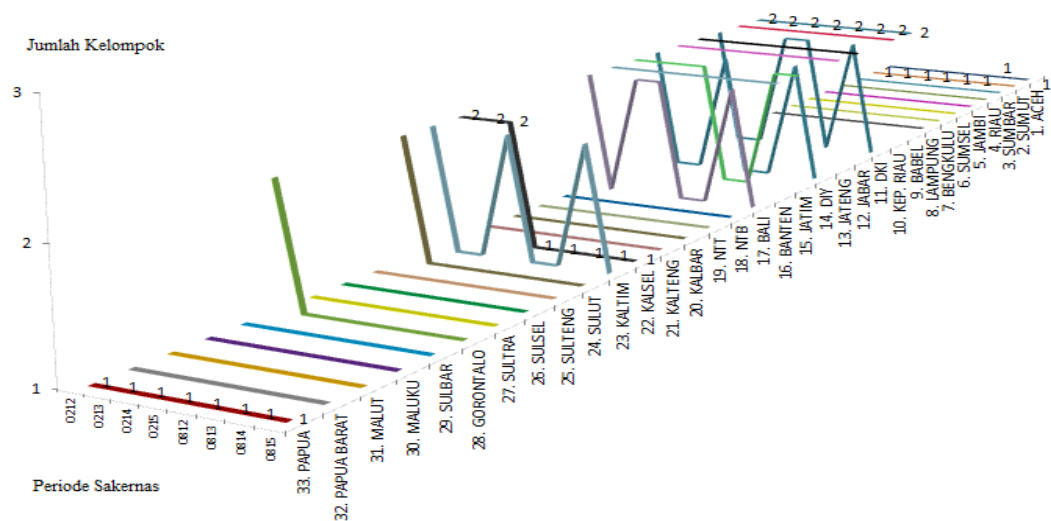
Lapangan usaha (sektor) pada analisis pasar tenaga kerja dibedakan menjadi 3 yaitu sektor pertanian (primer), sektor manufaktur (sekunder) dan sektor jasa-jasa (tersier). Jika pergeseran persentase penduduk yang bekerja dari sektor primer ke sektor sekunder dan tersier relatif lamban, itu pertanda bahwa kinerja perekonomian cenderung kurang dinamis atau minimal tidak berdampak pada pengembangan penyerapan tenaga kerja di sektor-sektor yang produktif (Hasbullah, 2013).



Gambar 4.3 Persentase Bekerja Menurut Lapangan Usaha, Agustus 2014-2015

Kondisi pasar tenaga kerja pada Agustus 2015 menunjukkan perubahan ke arah yang lebih baik. Persentase bekerja di sektor pertanian mengalami penurunan dibandingkan dengan tahun sebelumnya. Sementara sektor jasa-jasa dan sektor manufaktur menunjukkan peningkatan.

Gambar 4.4 menunjukkan beberapa provinsi konsisten berada pada kelompok yang sama dan beberapa provinsi berada pada kelompok yang berbeda-beda antar periode pencacahan Sakernas.



Gambar 4.4 *Plot Keanggotaan Cluster Subset Data k<sub>2</sub> 0212-k<sub>2</sub>0815*

Sekitar 60 persen provinsi di Indonesia konsisten berada di kelompok 1 selama 4 tahun, baik pada Sakernas Februari maupun Sakernas Agustus. Provinsi Bangka Belitung, Kepulauan Riau, Jawa Barat, Jawa Tengah, dan Banten adalah 5 provinsi yang konsisten berada di kelompok 2.

Tabel 4.9 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Lapangan Usaha, Agustus 2015

Kelompok 1		Kelompok 2
Aceh	Kalimantan Tengah	Kep. Bangka Belitung
Sumatera Utara	Kalimantan Selatan	Kep. Riau
Sumatera Barat	Kalimantan Timur	Jawa Barat
Riau	Sulawesi Utara	Jawa Tengah
Jambi	Sulawesi Tengah	Jawa Timur
Sumatera Selatan	Sulawesi Selatan	Banten
Bengkulu	Sulawesi Tenggara	
Lampung	Gorontalo	
DKI Jakarta	Sulawesi Barat	
DI Yogyakarta	Maluku	
Bali	Maluku Utara	
Nusa Tenggara Barat	Papua Barat	
Nusa Tenggara Timur	Papua	
Kalimantan Barat		

Pengelompokan provinsi di Indonesia berdasarkan subset  $k_2$  0815 menggunakan kelompok optimal model CCCC. Model CCCC memiliki nilai ICL terbesar pada jumlah kelompok  $G = 2$  yaitu 243,69 (**Lampiran 24**)

Hasil pengelompokan provinsi-provinsi di Indonesia berdasarkan persentase bekerja menurut lapangan usaha menunjukkan kelompok 1 merupakan provinsi dengan karakteristik penduduknya sebagian besar bekerja di sektor pertanian (43,08 persen) dan sedikit bekerja di sektor manufaktur (14,13 persen). Kelompok 2 merupakan provinsi dengan penduduk mayoritas bekerja di sektor jasa-jasa (47,33 persen), dan sedikit bekerja di sektor pertanian (23,73 persen). Sebaran pengelompokan provinsi di Indonesia pada Agustus 2015 juga dapat dilihat pada peta berikut ini.



Gambar 4.5 Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Lapangan Usaha ( $k_2$ ), Agustus 2015

Provinsi Bangka Belitung dan Kepulauan Riau merupakan provinsi di luar Jawa yang menjadi anggota kelompok 2. Semua provinsi di Pulau Jawa dan Bali pada Agustus 2015 tergabung di kelompok 2, kecuali DKI Jakarta dan DI Yogyakarta. Keanggotaan kelompok kedua provinsi ini cenderung berubah-ubah antar periode pencacahan Sakernas (**Lampiran 23**). DKI Jakarta memiliki karakteristik persentase pekerja sektor pertanian sangat rendah (0,42 persen), persentase pekerja sektor manufaktur di bawah rata-rata kelompok 2 yaitu 19,89 persen dan persentase pekerja di sektor jasa-jasa paling tinggi se-Indonesia yaitu 79,68 persen. Sedangkan DI Yogyakarta memiliki karakteristik tenaga kerja dengan persentase pekerja sektor pertanian sebesar 23,08 persen (di sekitar rata-

rata kelompok 2), persentase pekerja sektor manufaktur 23,77 persen (di sekitar kelompok 1), dan persentase pekerja sektor jasa-jasa cukup tinggi yaitu 53,15 persen.

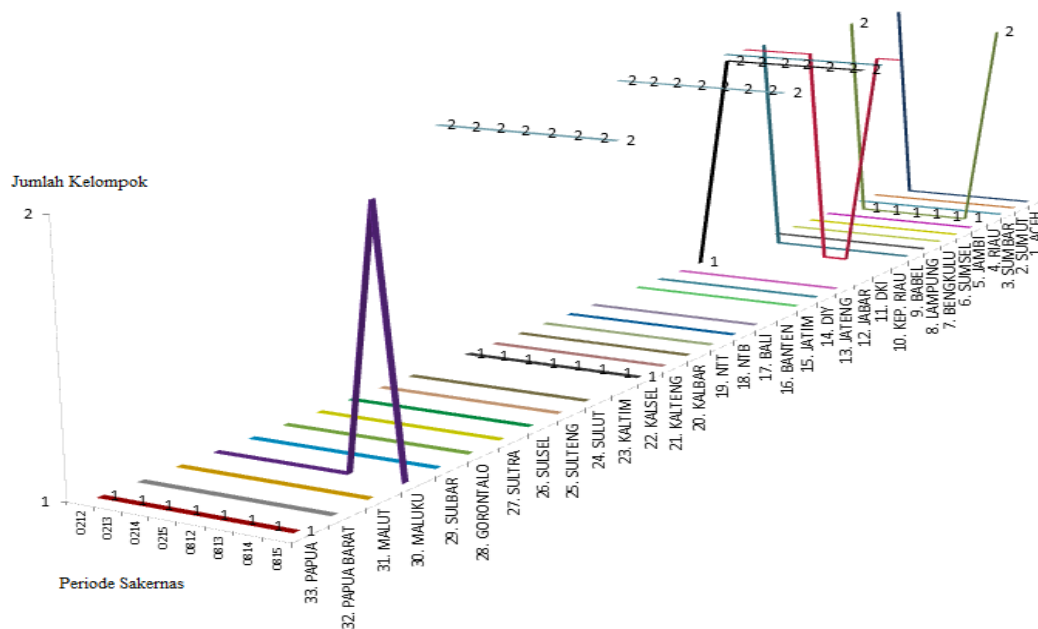
Tabel 4.10 Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Lapangan Usaha, Agustus 2015

Variable	k <sub>2</sub> 0815	Total Count	Mean	SE Mean	Coef Var	Min	Max
X <sub>9</sub>	1	27	43,08	14,71	34,15	0,42	73,93
	2	6	23,73	11,70	49,30	11,03	36,63
X <sub>10</sub>	1	27	14,13	4,46	31,59	4,59	23,77
	2	6	28,94	5,07	17,50	22,54	35,00
X <sub>11</sub>	1	27	42,79	11,13	26,00	21,47	79,68
	2	6	47,33	7,18	15,17	40,40	55,09

Dari Tabel 4.10 di atas diketahui jika *range* nilai kelompok 1 masih terlalu lebar dibandingkan dengan kelompok 2. Ini berarti di kelompok 1 terdapat provinsi yang karakteristiknya ekstrim berbeda dengan provinsi yang lain. Perlu kajian lebih lanjut supaya hasil pengelompokan dengan subset data k<sub>2</sub> 0815 tidak bias.

#### 4.5.2 Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan ( $k_3$ )

Bekerja menurut status dibagi menjadi enam kategori yaitu status karyawan ( $X_3$ ), pengusaha ( $X_4$ ), berusaha sendiri atau dibantu buruh tidak tetap/tidak dibayar ( $X_5$ ), pekerja bebas ( $X_6$ ) dan pekerja keluarga/tak dibayar ( $X_7$ ). Sedangkan TPT menurut pendidikan dibedakan atas 3 kategori yaitu TPT pendidikan dasar ke bawah ( $X_{20}$ ), pendidikan menengah ( $X_{21}$ ) dan pendidikan tinggi ( $X_{22}$ ). Pengelompokan menggunakan 8 indikator variabel KILM ini menghasilkan keanggotaan kelompok yang stabil selama tahun 2012-2015, baik pada Sakernas Februari maupun Sakernas Agustus.



Gambar 4.6 *Plot Keanggotaan Cluster Subset Data  $k_3$  0212- $k_3$  0815*

Lebih dari 80 persen provinsi di Indonesia konsisten berada di kelompok 1, baik pada Sakernas Februari maupun Sakernas Agustus. DKI Jakarta, Banten dan Kalimantan Timur adalah provinsi-provinsi yang selalu konsisten berada di Kelompok 2. Dari **Tabel 4.2** sebelumnya diketahui bahwa MBC-ICL menghasilkan kelompok optimal satu kelompok ( $G = 1$ ) pada subset data  $k_3$  0815. Oleh karena itu, pengelompokan provinsi di Indonesia berdasarkan kombinasi variabel  $k_3$  bulan Agustus 2015 berasal dari hasil pengelompokan jumlah kelompok *fixed*  $G = 2$  model CCCC. Model CCCC memiliki nilai ICL terbesar pada jumlah kelompok *fixed*  $G = 2$  yaitu -257,05 (**Lampiran 26**).

Pengelompokan provinsi di Indonesia berdasarkan subset data  $k_3$  0815 menghasilkan pengelompokan dimana kelompok 1 terdiri dari 27 provinsi dan kelompok 2 beranggotakan 6 provinsi. Hasil pengelompokan dapat dilihat pada Tabel 4.11 berikut.

Tabel 4.11 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Persentase Bekerja Menurut Status, dan TPT Menurut Pendidikan, Agustus 2015

Kelompok 1		Kelompok 2
Aceh	Kalimantan Barat	Riau
Sumatera Utara	Kalimantan Tengah	Kep. Riau
Sumatera Barat	Kalimantan Selatan	DKI Jakarta
Jambi	Sulawesi Utara	Jawa Barat
Sumatera Selatan	Sulawesi Tengah	Banten
Bengkulu	Sulawesi Selatan	Kalimantan Timur
Lampung	Sulawesi Tenggara	
Kep. Bangka Belitung	Gorontalo	
Jawa Tengah	Sulawesi Barat	
DI Yogyakarta	Maluku	
Jawa Timur	Maluku Utara	
Bali	Papua Barat	
Nusa Tenggara Barat	Papua	
Nusa Tenggara Timur		

Kelompok 1 merupakan provinsi dengan karakteristik dimana sebagian besar penduduknya bekerja dengan status berusaha sendiri atau dibantu buruh tidak dibayar (37 persen). Penduduk bekerja sebagai karyawan sebanyak 32 persen. Pekerja keluarga di kelompok 1 lebih banyak dibanding kelompok 2 yaitu 18,7 persen. Tingkat pengangguran terbuka pekerja dengan pendidikan tinggi di kelompok 1 juga lebih besar dibandingkan kelompok 2 yaitu 7,5 persen. Ini berarti angkatan kerja pendidikan tinggi di kelompok 1 lebih sulit mendapatkan pekerjaan yang diinginkan dibandingkan di kelompok 2. Kelompok 2 adalah provinsi-provinsi dimana sebagian besar penduduk bekerja sebagai karyawan yaitu 57,3 persen. Persentase pekerja yang berusaha sendiri atau dibantu buruh tidak tetap/ tidak dibayar menempati posisi kedua yaitu sebanyak 24,55 persen. Di kelompok 2 sedikit sekali pekerja berstatus pekerja bebas dan pekerja keluarga.



Angkatan kerja berpendidikan dasar dan menengah lebih sulit mendapatkan pekerjaan di kelompok 2. Hal ini ditunjukkan dengan nilai TPT pendidikan dasar dan menengah di kelompok 2 lebih besar dibanding kelompok 1. Sebaliknya, TPT pendidikan tinggi di kelompok 2 lebih rendah dibanding kelompok 1. Sebaran pengelompokan provinsi di Indonesia pada Agustus 2015 dapat dilihat pada peta berikut ini.



Gambar 4.7 Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan ( $k_3$ ), Agustus 2015

Jika dilihat dari sisi ekonomi, diketahui provinsi-provinsi yang tergabung di kelompok 2 merupakan provinsi dengan sentra ekonomi pertambangan, industri, dan jasa-jasa. Di kelompok 2 terdapat banyak perusahaan-perusahaan sehingga mampu menyerap tenaga kerja. Sedangkan Kelompok 1 merupakan provinsi dengan sentra ekonomi yang tidak didominasi oleh industri ataupun pertambangan. Jumlah pekerja rentan di kelompok 1 lebih besar dibanding kelompok 2. Pekerja rentan adalah pekerja berstatus selain karyawan dan pengusaha seperti berusaha sendiri, berusaha dibantu buruh tidak bayar, pekerja keluarga, dan pekerja bebas. Semakin besar persentase pekerja rentan, semakin rendah tingkat kesejahteraan pekerja. Hasil pengelompokan menunjukkan sebagian besar provinsi di Indonesia berada di kelompok 1. Ini berarti mayoritas pekerja di Indonesia memiliki tingkat kesejahteraan yang rendah.

Tabel 4.12 Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan, Agustus 2015

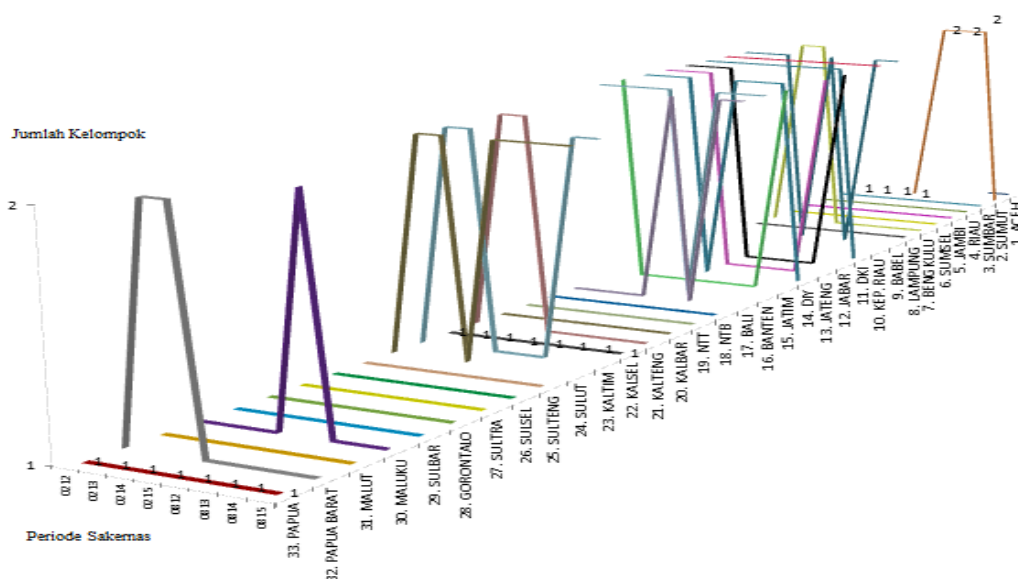
Variable	K <sub>3</sub> 0815	Total Count	Mean	SE Mean	Var	Min	Max
X <sub>3</sub>	1	27	32,27	1,49	53,54	17,56	45,31
	2	6	57,30	4,04	97,99	46,24	68,49
X <sub>4</sub>	1	27	3,28	0,18	0,74	0,95	4,64
	2	6	3,97	0,36	0,77	2,88	5,07
X <sub>5</sub>	1	27	37,34	0,82	16,27	29,19	44,64
	2	6	24,55	1,58	14,92	19,66	28,84
X <sub>6</sub>	1	27	8,35	0,94	21,06	1,44	19,91
	2	6	7,60	1,90	21,60	2,65	14,64
X <sub>7</sub>	1	27	18,77	1,31	41,11	8,74	37,14
	2	6	6,58	0,86	4,42	4,12	9,99
X <sub>20</sub>	1	27	2,65	0,19	0,82	0,82	4,90
	2	6	6,02	0,57	1,97	4,80	8,50
X <sub>21</sub>	1	27	10,43	0,66	10,56	3,15	18,11
	2	6	11,10	0,91	4,93	7,82	14,07
X <sub>22</sub>	1	27	7,48	0,50	6,06	3,79	14,92
	2	6	5,74	0,81	3,94	3,83	9,51

Berdasarkan **Tabel 4.12** dapat diketahui karakteristik pasar tenaga kerja antara kelompok 1 dan 2 jelas berbeda. Meskipun bukan kelompok optimal, hasil pengelompokan dengan jumlah kelompok *fixed* bisa menghasilkan pengelompokan yang baik. Hal ini terlihat dari keanggotaan cluster yang cenderung stabil antar periode Sakernas seperti ditunjukkan oleh **Gambar 4.6** sebelumnya.

#### 4.5.3 Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur Menurut Pendidikan ( $k_4$ )

Pada penelitian ini, analisis pasar tenaga kerja menurut jam kerja dibagi menjadi lima kategori yaitu Persentase Bekerja 1-14 Jam ( $X_{13}$ ), Persentase Bekerja 15-24 Jam ( $X_{14}$ ), Persentase Bekerja 25-34 Jam ( $X_{15}$ ), Persentase Bekerja 35-39 Jam ( $X_{16}$ ), dan Persentase Bekerja 40-48 Jam ( $X_{17}$ ). Sedangkan persentase pekerja setengah penganggur menurut pendidikan dibedakan atas 3 kategori yaitu setengah penganggur pendidikan dasar ke bawah ( $X_{20}$ ), setengah penganggur pendidikan menengah ( $X_{21}$ ) dan setengah penganggur pendidikan tinggi ( $X_{22}$ ).

Batasan waktu seseorang disebut bekerja adalah satu jam dalam seminggu. Seseorang akan dianggap sebagai bekerja jika dia melakukan pekerjaan yang mendapatkan imbalan ekonomis minimal selama satu jam dalam seminggu dan dilakukan secara berturut turut. Perlu dipahami bahwa konsep bekerja satu jam dalam seminggu yang dipakai dalam pendataan adalah untuk menjangkau semua orang Indonesia yang bekerja walau pekerjaannya hanya terbatas dari sisi penggunaan waktu. Dan ternyata, walaupun jumlahnya sangat kecil, tetapi ada angkatan kerja di Indonesia yang bekerja dengan kategori tersebut. Akan tetapi, indikator persentase bekerja 1 jam dan lebih dari 48 jam tidak disertakan pada penelitian ini.



Gambar 4.8 *Plot Keanggotaan Cluster Subset Data  $k_4$  0212- $k_4$  0815*

Sekitar 50 persen provinsi di Indonesia konsisten berada di kelompok 1, baik pada Sakernas Februari maupun Sakernas Agustus. Kepulauan Riau adalah satu-satunya provinsi yang selalu konsisten berada di Kelompok 2. Dari **Tabel 4.2** sebelumnya diketahui bahwa MBC-ICL menghasilkan kelompok optimal satu kelompok ( $G = 1$ ) pada subset data  $k_4$  0815. Oleh karena itu, pengelompokan provinsi di Indonesia berdasarkan kombinasi variabel  $k_4$  bulan Agustus 2015 berasal dari hasil pengelompokan jumlah kelompok *fixed*  $G = 2$  model CCCU. Model CCCU memiliki nilai ICL terbesar pada jumlah kelompok *fixed*  $G = 2$  yaitu -205,31 (Lampiran **28**). Pengelompokan provinsi di Indonesia berdasarkan subset data  $k_4$  0815 menghasilkan pengelompokan dimana kelompok 1 terdiri dari 24 provinsi dan kelompok 2 beranggotakan 9 provinsi. Hasil pengelompokan dapat dilihat pada Tabel **4.13** berikut.

Tabel 4.13 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur, Agustus 2015

Kelompok 1		Kelompok 2
Aceh	Kalimantan Barat	Kep. Bangka Belitung
Sumatera Utara	Kalimantan Tengah	Kep. Riau
Sumatera Barat	Kalimantan Selatan	Jawa Barat
Riau	Sulawesi Tengah	Jawa Tengah
Jambi	Sulawesi Selatan	Jawa Timur
Sumatera Selatan	Sulawesi Tenggara	Banten
Bengkulu	Gorontalo	Bali
Lampung	Sulawesi Barat	Kalimantan Timur
DKI Jakarta	Maluku	Sulawesi Utara
DI Yogyakarta	Maluku Utara	
Nusa Tenggara Barat	Papua Barat	
Nusa Tenggara Timur	Papua	

Kelompok 1 dan kelompok 2 memiliki karakteristik yang sama, dimana mayoritas pekerjaanya bekerja selama 40-48 jam ( $X_{17}$ ). Kelompok 1 sebesar 24,23 persen dan kelompok 2 sebesar 36,94 persen. Kedua kelompok ini juga mayoritas angkatan kerjanya setengah penganggur berpendidikan dasar. Kelompok 1 sebesar 63,88 persen dan kelompok 2 sebesar 69,52 persen. Jumlah pekerja yang bekerja

di bawah 40 jam ( $X_{13}$ ,  $X_{14}$ ,  $X_{15}$ , dan  $X_{16}$ ) di kelompok 1 selalu lebih besar dibandingkan dengan kelompok 2. Begitu pula untuk angkatan kerja setengah penganggur pendidikan menengah dan pendidikan atas di kelompok 1 selalu lebih tinggi dibandingkan dengan kelompok 2. Sebaran pengelompokan provinsi di Indonesia pada Agustus 2015 dapat dilihat pada peta berikut ini.



Gambar 4.9 Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur, Agustus 2015

Semua provinsi di Pulau Jawa dan Bali pada Agustus 2015 tergabung di kelompok 2, kecuali DKI Jakarta dan DI Yogyakarta. Padahal kedua provinsi ini pada tahun-tahun sebelumnya berada di kelompok 2 (**Lampiran 27**). Pada Agustus 2015 bergeser ke kelompok 1. Persentase pekerja setengah penganggur pendidikan dasar ( $X_{23}$ ) DKI Jakarta dan DIY Yogyakarta pada agustus 2015 mengalami penurunan dibandingkan dengan agustus 2014. Nilainya berada di luar *range* kelompok 2. Akan tetapi, untuk variabel indikator yang lain, kedua provinsi ini masih menunjukkan kecenderungan berada pada *range* kelompok 2 dibandingkan kelompok 1. Oleh karena itu tidak mengherankan jika *range* nilai kelompok 1 masih terlalu lebar dibandingkan dengan kelompok 2. Ini berarti di kelompok 1 terdapat provinsi yang karakteristiknya ekstrim berbeda dengan provinsi yang lain. Perlu kajian lebih lanjut supaya hasil pengelompokan dengan subset data  $k_4$  0815 tidak bias.

Tabel 4.14 Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Bekerja Menurut Jam Kerja, dan Pekerja Setengah Penganggur, Agustus 2015

Variable	K <sub>3</sub> 0815	Total Count	Mean	SE Mean	Coef Var	Min	Max
X <sub>13</sub>	1	24	6,71	0,58	41,95	2,52	15,93
	2	9	3,99	0,48	35,70	2,03	6,74
X <sub>14</sub>	1	24	13,84	0,68	24,23	3,51	20,31
	2	9	8,72	0,75	25,67	5,81	12,22
X <sub>15</sub>	1	24	15,84	0,84	25,95	4,60	27,05
	2	9	10,60	0,75	21,13	6,43	13,47
X <sub>16</sub>	1	24	11,84	0,55	23,01	5,59	21,67
	2	9	9,04	0,42	13,70	7,22	10,61
X <sub>17</sub>	1	24	24,23	1,84	37,27	2,46	48,96
	2	9	36,94	1,91	15,53	30,14	47,26
X <sub>23</sub>	1	24	63,88	2,06	15,77	36,87	83,87
	2	9	69,52	3,06	13,20	57,04	79,71
X <sub>24</sub>	1	24	27,04	1,35	24,38	14,24	41,03
	2	9	24,38	3,10	38,18	13,55	39,72
X <sub>25</sub>	1	24	9,08	0,96	51,90	1,88	22,09
	2	9	6,11	0,72	35,27	3,24	9,53

Dari Tabel 4.14 diketahui lebih dari separuh angkatan kerja Indonesia adalah pekerja setengah penganggur dengan pendidikan dasar ke bawah, baik pada kelompok 1 (63,88 persen) maupun kelompok 2 (69,52 persen). Ini berarti, pekerja di Indonesia sebagian besar bekerja kurang dari 35 jam dalam seminggu yang lalu, dan masih belum puas dengan pekerjaan yang dimiliki saat ini. Hal ini ditunjukkan dengan kesediaan mereka menerima pekerjaan tambahan atau mencari pekerjaan lain dengan jumlah jam kerja yang lebih banyak.

#### 4.5.4 Karakteristik Pasar Tenaga Kerja Indonesia Berdasarkan Indikator EPR, Pekerja Rentan, dan Pekerja Sektor Informal (k<sub>5</sub>)

*Employment to Population Ratio* (EPR) didefinisikan sebagai proporsi penduduk umur kerja suatu negara yang berstatus bekerja terhadap penduduk umur kerja. Rasio yang tinggi berarti sebagian besar penduduk suatu negara adalah bekerja. Sementara rasio rendah berarti bahwa sebagian besar penduduk tidak terlibat langsung dalam kegiatan yang berhubungan dengan pasar tenaga kerja. Rasio ini memberikan informasi tentang kemampuan ekonomi untuk menciptakan lapangan kerja. Di banyak negara, indikator ini menghasilkan analisis yang lebih mendalam dibandingkan dengan tingkat pengangguran. Pekerja rentan (*vulnerable employment*) adalah pekerja yang mencakup pekerja dengan status berusaha sendiri, berusaha dibantu buruh tidak tetap/ tak dibayar, pekerja bebas dan pekerja keluarga. Sedangkan pekerja sektor informal adalah angkatan kerja yang bekerja di unit produksi dalam usaha rumah tangga yang dimiliki oleh rumah tangga. Sebaran pengelompokan provinsi di Indonesia pada Agustus 2015 dapat dilihat pada peta berikut ini.



Gambar 4.10 Peta Pengelompokan Provinsi di Indonesia Berdasarkan Indikator EPR, Pekerja Rentan, dan Pekerja Sektor Informal, Agustus 2015

RMBC-MML mengelompokkan provinsi di Indonesia berdasarkan ketiga indikator ini pada bulan Agustus 2015 ke dalam 2 kelompok. Dimana kelompok 1 terdiri dari 12 provinsi dan kelompok 2 beranggotakan 21 provinsi.

Tabel 4.15 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator EPR ( $X_2$ ), Pekerja Rentan ( $X_8$ ), dan Pekerja Sektor Informal ( $X_{18}$ ), Agustus 2015

Kelompok 1		Kelompok 2	
1	Sumatera Utara	1	Aceh
2	Jambi	2	Sumatera Barat
3	Sumatera Selatan	3	Riau
4	Bengkulu	4	Lampung
5	Jawa Barat	5	Kep. Bangka Belitung
6	Nusa Tenggara Timur	6	Kep. Riau
7	Kalimantan Barat	7	DKI Jakarta
8	Gorontalo	8	Jawa Tengah
9	Sulawesi Barat	9	DI Yogyakarta
10	Maluku Utara	10	Jawa Timur
11	Papua Barat	11	Banten
12	Papua	12	Bali
		13	Nusa Tenggara Barat
		14	Kalimantan Tengah
		15	Kalimantan Selatan
		16	Kalimantan Timur
		17	Sulawesi Utara
		18	Sulawesi Tengah
		19	Sulawesi Selatan
		20	Sulawesi Tenggara
		21	Maluku

Secara rata-rata, kelompok 1 dan kelompok 2 memiliki nilai EPR yang hampir sama yaitu di atas 60 persen. Ini berarti, lebih dari 60 persen angkatan kerja di kedua kelompok ini sudah memiliki pekerjaan.

Jawa Barat merupakan satu-satunya provinsi di Pulau Jawa yang tergabung di kelompok 1 pada Agustus 2015. Jawa barat memiliki angka EPR terendah di Indonesia yaitu 55,08 persen. Sementara itu, persentase pekerja rentan di provinsi ini cukup tinggi yaitu 50,39 persen. Kelompok 1 teridentifikasi memiliki persentase pekerja rentan dan pekerja yang bekerja di sektor informal lebih tinggi dibandingkan kelompok 2. Papua terdeteksi sebagai provinsi dengan nilai EPR dan persentase pekerja rentan tertinggi di Indonesia yaitu masing-masing 76,49 persen dan 81,49 persen. Papua juga merupakan provinsi dengan



persentase pekerja sektor informal terbesar yaitu 79,96 persen. Kondisi ini sangat perlu menjadi perhatian. Besarnya penciptaan lapangan kerja di Papua tidak serta-merta meningkatkan kesejahteraan rakyatnya. Karena sebagian besar pekerja di Papua adalah pekerja dengan pendapatan rendah yang dicirikan dengan besarnya persentase pekerja rentan dan pekerja di sektor informal.

Di kelompok 2, Bali merupakan provinsi dengan nilai EPR tertinggi. 74,01 persen angkatan kerja di provinsi ini memiliki pekerjaan. Berbeda dengan Papua, meskipun memiliki nilai EPR tertinggi, persentase pekerja rentan dan pekerja sektor informal di Provinsi Bali berada di bawah rata-rata kelompok 2. Ini berarti sebagian besar pekerja di Bali sudah memperoleh pekerjaan layak. Provinsi Nusa Tenggara Barat merupakan provinsi yang teridentifikasi sebagai provinsi dengan persentase pekerja rentan dan pekerja sektor informal tertinggi di kelompok 2 yaitu 73,2 persen dan 65,67 persen.

Tabel 4.16 Statistik Deskriptif Pengelompokan Provinsi di Indonesia Berdasarkan Indikator EPR, Pekerja Rentan, dan Pekerja Sektor Informal, Agustus 2015

Variable	K <sub>5</sub> 0815	Total Count	Mean	SE Mean	Coef Var	Min	Max
X <sub>2</sub>	1	12	64,66	1,46	7,82	55,08	76,40
	2	21	62,05	0,98	7,29	55,75	74,01
X <sub>8</sub>	1	12	65,89	2,56	13,47	50,39	81,49
	2	21	55,89	2,88	23,61	27,04	73,20
X <sub>18</sub>	1	12	61,51	2,82	15,89	43,80	79,96
	2	21	50,10	2,79	25,52	21,49	65,67

Dari tabel statistik deskriptif di atas dapat dilihat kelompok 1 dan kelompok 2 memiliki nilai EPR yang hampir sama. Akan tetapi, pekerja di kelompok 1 lebih banyak yang berstatus sebagai pekerja rentan dan pekerja sektor informal dibandingkan dengan kelompok 2.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan, dapat diambil beberapa kesimpulan penelitian ditinjau menurut metode dan kasus sesuai dengan permasalahan yang dirumuskan pada bab 1 sebelumnya.

1. Subset data KILM yang memenuhi asumsi distribusit  $t$  multivariat adalah subset data yang dibentuk berdasarkan kombinasi variabel  $k_2$ ,  $k_3$ ,  $k_4$ , dan  $k_5$ .
  - a) Berdasarkan hasil uji asumsi distribusi dengan *mardia test*, subset data KILM yang terbentuk dari kombinasi variabel  $k_1$  berdistribusi normal multivariat. Sehingga subset data KILM kombinasi variabel  $k_1$  tidak tepat diaplikasikan pada *model-based clustering mixture t* multivariat.
  - b) Berdasarkan hasil deteksi *outlier* menggunakan *dd plot* diketahui 40 subset data KILM yang digunakan pada penelitian ini mengandung *outlier* (kuantil 90 persen). Ini berarti asumsi bahwa data indikator pasar tenaga kerja Indonesia cenderung memuat provinsi *outlier* dengan data yang berbeda ekstrim dibanding provinsi lain telah terbukti.
2. Dipilih kriteria *Integrated Completed Likelihood* (ICL) untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator pasar tenaga kerja.
  - a) Pengelompokan dengan MBC-ICL menghasilkan kelompok optimal yang konsisten pada setiap ulangan. Kekonsistenan hasil pengelompokan tidak hanya dari jumlah kelompok optimal tetapi juga anggota dalam kelompoknya. Antar ulangan selalu menghasilkan jumlah dan anggota kelompok optimal yang sama.
  - b) Pengelompokan dengan RMBC-MML berdasarkan subset data  $k_3$  0815 dan  $k_4$  0815 tidak berhasil membentuk kelompok atau terjadi *error* saat subset data diolah dengan program RMBC-MML.
  - c) Sebagian besar kriteria ICL menghasilkan jumlah kelompok optimal  $G = 1$  pada Sakernas Agustus 2012-2015. Oleh karena itu, untuk

menganalisis pasar tenaga kerja Indonesia Agustus 2015 dilakukan pengelompokan dengan jumlah kelompok yang ditetapkan (*fixed*) dan dipilih model dengan nilai ICL paling besar.

- d) Kelompok optimal RMBC-MML cocok digunakan untuk analisis pasar tenaga kerja Indonesia berdasarkan subset data  $k_5$  0815.
3. Pengelompokan provinsi di Indonesia berdasarkan Indikator Bekerja Menurut Status, dan TPT Menurut Pendidikan (kombinasi variabel  $k_3$ ) menghasilkan pengelompokan yang lebih stabil selama 4 tahun periode pencacahan Sakernas 2012-2015, dibandingkan dengan kombinasi variabel  $k_2$ ,  $k_4$  dan  $k_5$ .
  - a. Berdasarkan pengelompokan menggunakan subset data  $k_2$  0815 diketahui bahwa kondisi pasar tenaga kerja Indonesia pada Agustus 2015 menunjukkan perubahan ke arah yang lebih baik. Persentase penduduk yang bekerja di sektor pertanian mengalami penurunan dibandingkan dengan tahun sebelumnya. Sementara persentase penduduk yang bekerja di sektor jasa-jasa dan sektor manufaktur menunjukkan peningkatan.
  - b. Berdasarkan pengelompokan menggunakan subset data  $k_3$  0815 diketahui lebih dari 80 persen provinsi di Indonesia konsisten berada di kelompok 1, yaitu kelompok yang memiliki karakteristik sebagian besar penduduknya bekerja dengan status berusaha sendiri atau dibantu buruh tidak dibayar.
  - c. Berdasarkan pengelompokan menggunakan subset data  $k_4$  0815 diketahui lebih dari separuh angkatan kerja Indonesia adalah pekerja setengah penganggur dengan pendidikan dasar ke bawah, baik pada kelompok 1 (63,88 persen) maupun kelompok 2 (69,52 persen).
  - d. Berdasarkan pengelompokan menggunakan subset data  $k_5$  0815 diketahui Provinsi Papua teridentifikasi sebagai provinsi dengan persentase pekerja rentan dan persentase pekerja sektor informal tertinggi di Indonesia.

## 5.2 Saran

1. Penelitian ini tidak mengelompokkan provinsi di Indonesia berdasarkan kombinasi variabel  $k_1$  (indikator TPAK, Pekerja paruh waktu, Pekerja sektor informal, dan TPT) dikarenakan subset data  $k_1$  berdistribusi normal multivariat. Penelitian selanjutnya bisa melakukan pengelompokan provinsi di

Indonesia berdasarkan kombinasi variabel  $k_1$  menggunakan *model-based clustering* normal multivariat (Banfield dan Raftery , 1993).

2. Pada penelitian selanjutnya, perlu dikembangkan pengelompokan provinsi di Indonesia berdasarkan indikator pasar tenaga kerja menggunakan kriteria pemilihan model *clustering* lainnya, sehingga diperoleh hasil pengelompokan yang lebih stabil antar periode pencacahan Sakernas.
3. Pada penelitian selanjutnya, bisa dilakukan pengelompokan kabupaten/kota di Indonesia (Siagian, 2014) menurut indikator pasar tenaga kerja menggunakan RMBC dengan kriteria MML.
4. Perlu dikembangkan penelitian pada program RMBC-MML agar bisa dihasilkan kelompok optimal pada subset data  $k_2$  dan  $k_3$ .
5. Pada penelitian selanjutnya, bisa dilakukan pengelompokan menggunakan kombinasi indikator KILM dengan indikator sosial ekonomi lainnya seperti persentase penduduk miskin, laju pertumbuhan dan sebagainya.
6. Kebijakan pemerintah terkait tenaga kerja sebaiknya difokuskan kepada provinsi-provinsi dengan persentase pekerja rentan dan pekerja sektor informal tinggi seperti Papua, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Barat, dan beberapa provinsi lainnya. Provinsi-provinsi ini juga sebaiknya menjadi tujuan utama program kerja layak (*decent work country programme*) oleh ILO.

*Halaman ini sengaja dikosongkan*

## DAFTAR PUSTAKA

- Andrews, J.L., McNicholas, P.D., Subedi, S. (2011). Model-based classification via mixtures of multivariate t-distributions. *Computational Statistics and Data Analysis*, 2.
- Banfield, J.D. & Raftery, A.E . (1993). Model-Based Gaussian and non-Gaussian Clustering. *Biometrics Vol.49, No.3*, hal 803-821.
- Baudry, J.P., Cardoso, M., Celeux, G., Amorim, M.J., Ferreira, A.S. (2013). Enhancing the selection of a model-based clustering with external categorical variables. *Advances in Data Analysis and Classification*, hal1-20, Springer Berlin Heidelberg. ISSN/ISSBN; 1862-5347.
- Bezdek, J.C., Ehrlich, R., Full, W.E. (1984). FCM : The Fuzzy C-Means Clustering Algorithm. *Computers & Geosciences.10(2-3)*, 191-203.
- BI. (2015). Survei Konsumen.
- Biernacki, C., Celeux, G. & Govaert, G. (2000). Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE Trans. Pattern Anal. Mach. Intel.*, 719-725.
- Bouveyron, C. & Brunet-Saumard, C. (2014). Model-based clustering of high dimensional data: A review. *Computational Statistics and Data Analysis, Journal of the American Statistical Association*, 71, 52-78.
- BPS. (2012a). Indikator Pasar Tenaga Kerja Indonesia Februari 2012. Jakarta: Badan Pusat Statistik.
- BPS. (2012b). Indikator Pasar Tenaga Kerja Indonesia Agustus 2012. Jakarta: Badan Pusat Statistik.
- BPS. (2013a). Indikator Pasar Tenaga Kerja Indonesia Februari 2013. Jakarta: Badan Pusat Statistik.
- BPS. (2013b). Indikator Pasar Tenaga Kerja Indonesia Agustus 2013. Jakarta: Badan Pusat Statistik.
- BPS. (2014a). Indikator Pasar Tenaga Kerja Indonesia Februari 2014. Jakarta: Badan Pusat Statistik.

- BPS. (2014b). Indikator Pasar Tenaga Kerja Indonesia Agustus 2014. Jakarta: Badan Pusat Statistik.
- BPS. (2015a). Indikator Pasar Tenaga Kerja Indonesia Februari 2015. Jakarta: Badan Pusat Statistik.
- BPS. (2015b). Indikator Pasar Tenaga Kerja Indonesia Agustus 2015. Jakarta: Badan Pusat Statistik.
- BPS. (2015c). Statistik Indonesia. Jakarta: Badan Pusat Statistik.
- BPS. (2015d, Agustus). Sakernas.
- BPS. (2016, Mei 09). *Badan Pusat Statistik*. Dipetik Agustus 12, 2016, dari Badan Pusat Statistik: <http://www.bps.go.id>
- BPS. (2016, Februari). Sakernas.
- Casella, G. & Berger, R.L. (2002). *Statistical Inference, Second Editon*. Pacific Grove: Thomson Learning Inc.
- Celeux, G. & Govaert, G. (1995). Gaussian Parsimonious Clustering Models, Pattern Recognition. *Journal of the American Statistical Association*, Vol.28, hal 781-793.
- Cozzini, A., Asra, A., & Montana, G. (2013). Model-based clustering with gene ranking using penalized mixtures of heavy tailed distributions. *Journal of Bioinformatics and Computational Biology*, 11(3).
- Cuesta, J. A. & Albertos. (2008). Robust estimation in the normal mixture model based on robust clustering. *Journal of the Royal Statistical Society : Series B*, hal 779-802.
- Damayanti, F. (2015). Model based clustering mixture t-multivariat dengan kriteria Integrated classification likelihood (Pengelompokan Provinsi di Indonesia menurut Capaian Berkelanjutan Tahun 2011). *Tesis Institut Teknologi Sepuluh November*.
- Dasgupta, A. & Raftery, A. E. (1998). Detecting features in spatial point processes with cluster via model-based clustering. *Journal of the American Statistical Association*, Vol. 93, hal 294-302.
- Day, N. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, Vol.56, hal 463-474.

- Dekking, F.M., Kraaikamp, C., Lopuha, H.P., Mester, L.E. (2005). *A Modern introduction to Probability and Statistics*. London: Springer.
- Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum Likelihood for Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society, Series. B*, 39, 1-38.
- Edwards & Cavalli-Sforza. (1965). A method for cluster analysis. *Biometrics*, Vol.21, hal 362-375.
- Ehrenberg, R.G. & Smith, R.S. (2012). *Modern Labor Economics : Theory and Public Policy, Eleventh Ed*. Boston: Pearson.
- Figueiredo, M.A.T. & Jain, A. K. (2002). Unsupervised Learning of Finite Mixture Models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 29(3), 1-15.
- Forbes, C., Evans, M., Hastings, N. (2011). *Statistical Distributions, Fourth Edition*. New Jersey: John Wiley & Sons, Inc.
- Fraley, C. & Raftery, A.E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*, 41(8), 578-588.
- Fraley, C. &. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, Vol. 97, hal 611-631.
- Hammouda, K., Karray, F. (2000). A comparative study of data clustering techiques. . *Course Project SYDE 625* (hal. Tools of Intelligent Systems Design). Ontario: Univ. of Waterloo.
- Hasbullah, J. (2013). *Tangguh dengan Statistik dalam Membaca Realita Dunia*. Bandung: Nuansa Cendekia.
- Hogg, R.V. & Tanis, E.A. (1997). *Probability and Statistical Inference, 5th edition*. New Jersey: Prentice Hall Inc.
- ILO. (2011). *Decent Work Country Profile Indonesia*. Geneva: International Labour Organization.
- ILO. (2014a). *Key Indicators of The Labour Market, Eight Edition*. Geneva: International Labour Organization.



- ILO. (2014b). *Decent Work Country Profile Indonesia*. Geneva: International Labour Organization.
- ILO. (2016). *Indonesia Decent Work Country Programme 2012-2015*. Jakarta: International Labour Organization.
- Jain, A. (2010). Data clustering : 50 years beyond K-mean. *Pattern Recognition Letters* 31 (8), 651-666.
- Johnson, R. A. (2007). *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall.
- Johnson, R.A. & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis 6th edition*. New Jersey: Pearson Prentice Hall.
- Kass, R. E. & Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, hal 928-934.
- Kass, R.E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, Vol. 90, 773-795.
- Keribin, C. (2000). Consistent estimation of the order of mixture models. *The Indian Journal of Statistics, Series A. Vol. 62, No. 1*, hal. 49-66.
- Khattree, R., & Naik, D. N. (2000). *Multivariate data reduction and discrimination with SAS software*. Sas Institute.
- Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *The Annals of Statistics*, Vol. 20, hal. 1350-1360.
- Manning, C. (2001). Lesson From Labour Market Adjustment to the East Asean. *7th Convention of the East Asian Economic Association*, (hal. 9). Singapore.
- McLachlan, G. (1982). The classification and mixture maximum likelihood approaches to cluster analysis. *Volume 2 of Handbook of Statistics*, hal 199-208.
- McLachlan, G.J. & Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley and Sons.
- McLachlan, G.J. & Wang, S.K. (2012). Clustering of time-course gene expression profiles using normal mixture models with autoregressive random effects. *BMC Bioinformatics*, hal 300.

- McLachlan, G.J., Ng, S.K. & Bean, R.W. (2004). Robust Mixture Modeling. *Proceedings of Third International American Statistical Association, Physical and Engineering Sciences Section* (hal. 2044-2055). Toronto, Alexandria, Virginia: American Statistical Association.
- Melnykov, V. & Maitra, R. (2010). Finite Mixture Models and Model-Based Clustering. *Statistics Surveys*, 4, 80-116.
- Meng, X. & Rubin, D.B. (1993). Maximum Likelihood estimation via the ECM Algorithm : A General Framework. *Biometrika*, 80, 2, Oxford University Press.
- Pardede, T. (2008). Perbandingan metode berbasis model (model-based) dengan metode-metode K-Means dalam analisis gugus. *Jurnal Sigma, Sains dan Teknologi*, Vol. 11, No. 2, 157-166.
- Rencher, A. (2002). *Methods of Multivariate Analysis*. Canada: John Wiley & Sons, Inc.,.
- Rousseeuw, P.J, & Van Zomeren, B.C . (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association* Vol 85, 633-639.
- Rousseeuw, P.J. & Van Driessen, K. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41, 212-223.
- Saliman. (2005). Dampak Krisis Terhadap Ketenagakerjaan Indonesia. *Jurnal Ekonomi dan Pendidikan Universitas Negeri Yogyakarta*, Vol 2 No 1.
- Sharma, S. (1996). *Applied Multivariate Techniques*. Canada: John Wiley & Sons, Inc.
- Siagian, T. H. (2014). Robust Model-Based Clustering dengan Distribusi t Multivariat dan Minimum Message Length. *Disertasi FMIPA Institut Teknologi Sepuluh Nopember*.
- Titterington, D.M., A. F. M. Smith & U. E. Makov. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: John Wiley & Sons.
- Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distribution*. Ltd: John Wiley & Sons.
- Wallace, C.S. & Boulton, D.M. (1968). An Information Measure for Classification. *Computer Journal*, 11, hal 185-194.

- Wallace, C.S. & Freeman, P.R. (1987). Estimation and Inference by Compact Coding. *Journal of the Royal Statistical Society B*, 49(3), hal 240-265.
- Wolfe, J. (1965). A computer program for the maximum-likelihood analysis of types. *USNPRA Technical Bulletin*, hal 65-15.

Lampiran 1. Tata Nama *Model-Based Clustering* dalam *Package Teigen Software R*

No	Model	$\lambda_g$	$D_g$	$A_g$	$v_g$
1	UUUU	U	U	U	U
2	UUUC	U	U	U	C
3	CUCU	C	U	C	U
4	CUCC	C	U	C	C
5	CUUU	C	U	U	U
6	CUUC	C	U	U	C
7	CCCU	C	C	C	U
8	CCCC	C	C	C	C
9	CIUU	C	I	U	U
10	CIUC	C	I	U	C
11	CICU	C	I	C	U
12	CICC	C	I	C	C
13	UIIU	U	I	I	U
14	UIIC	U	I	I	C
15	CIIU	C	I	I	U
16	CIIC	C	I	I	C
17	UIUU	U	I	U	U
18	UIUC	U	I	U	C
19	UCCU	U	C	C	U
20	UCCC	U	C	C	C
21	UUCU	U	U	C	U
22	UUCC	U	U	C	C
23	UICU	U	I	C	U
24	UICC	U	I	C	C
25	UCUU	U	C	U	U
26	UCUC	U	C	U	C
27	CCUU	C	C	U	U
28	CCUC	C	C	U	C

Keterangan :

C : terbatas (*constrained*), artinya antar *cluster* sama

U : tidak ada batasan (*unconstrained*), artinya antar *cluster* berbeda

I : matriks Identitas

Lampiran 2. Tata Nama Kombinasi Variabel Subset Data KILM

Kombinasi	<u>Sakernas Februari</u>				<u>Sakernas Agustus</u>			
	2012	2013	2014	2015	2012	2013	2014	2015
k <sub>1</sub>	k <sub>1</sub> 0212	k <sub>1</sub> 0213	k <sub>1</sub> 0214	k <sub>1</sub> 0215	k <sub>1</sub> 0812	k <sub>1</sub> 0813	k <sub>1</sub> 0814	k <sub>1</sub> 0815
k <sub>2</sub>	k <sub>2</sub> 0212	k <sub>2</sub> 0213	k <sub>2</sub> 0214	k <sub>2</sub> 0215	k <sub>2</sub> 0812	k <sub>2</sub> 0813	k <sub>2</sub> 0814	k <sub>2</sub> 0815
k <sub>3</sub>	k <sub>3</sub> 0212	k <sub>3</sub> 0213	k <sub>3</sub> 0214	k <sub>3</sub> 0215	k <sub>3</sub> 0812	k <sub>3</sub> 0813	k <sub>3</sub> 0814	k <sub>3</sub> 0815
k <sub>4</sub>	k <sub>4</sub> 0212	k <sub>4</sub> 0213	k <sub>4</sub> 0214	k <sub>4</sub> 0215	k <sub>4</sub> 0812	k <sub>4</sub> 0813	k <sub>4</sub> 0814	k <sub>4</sub> 0815
k <sub>5</sub>	k <sub>5</sub> 0212	k <sub>5</sub> 0213	k <sub>5</sub> 0214	k <sub>5</sub> 0215	k <sub>5</sub> 0812	k <sub>5</sub> 0813	k <sub>5</sub> 0814	k <sub>5</sub> 0815

Keterangan :

k<sub>1</sub> 0212 = subset data kombinasi variabel k<sub>1</sub> berdasarkan data Sakernas Februari 2012

k<sub>1</sub> 0812 = subset data kombinasi variabel k<sub>1</sub> berdasarkan data Sakernas Agustus 2012

### Lampiran 3. Variabel Penelitian

<b>Variabel</b>	<b>Keterangan</b>
X <sub>1</sub>	Tingkat Partisipasi Angkatan Kerja (TPAK)
X <sub>2</sub>	Rasio Penduduk yang Bekerja Terhadap Jumlah Penduduk (EPR)
X <sub>3</sub>	Persentase Bekerja Status Karyawan
X <sub>4</sub>	Persentase Bekerja Status Pengusaha
X <sub>5</sub>	Persentase Bekerja Status Berusaha Sendiri atau Dibantu Buruh Tidak Tetap/ Tidak Dibayar
X <sub>6</sub>	Persentase Bekerja Status Pekerja Bebas
X <sub>7</sub>	Persentase Bekerja Status Pekerja Keluarga/tak dibayar
X <sub>8</sub>	Persentase Bekerja Status Pekerja Rentan
X <sub>9</sub>	Persentase Bekerja di Sektor Pertanian
X <sub>10</sub>	Persentase Bekerja di Sektor Manufaktur
X <sub>11</sub>	Persentase Bekerja di Sektor Jasa-Jasa
X <sub>12</sub>	Tingkat Pekerja Paruh Waktu
X <sub>13</sub>	Persentase Bekerja 1-14 Jam
X <sub>14</sub>	Persentase Bekerja 15-24 Jam
X <sub>15</sub>	Persentase Bekerja 25-34 Jam
X <sub>16</sub>	Persentase Bekerja 35-39 Jam
X <sub>17</sub>	Persentase Bekerja 40-48 Jam
X <sub>18</sub>	Persentase Penduduk Bekerja di Sektor Informal
X <sub>19</sub>	Tingkat Pengangguran Terbuka-TPT
X <sub>20</sub>	TPT Pendidikan Dasar ke Bawah
X <sub>21</sub>	TPT Pendidikan Menengah
X <sub>22</sub>	TPT Pendidikan Tinggi
X <sub>23</sub>	Persentase Setengah Penganggur Pendidikan Dasar ke Bawah
X <sub>24</sub>	Persentase Setengah Penganggur Pendidikan Menengah
X <sub>25</sub>	Persentase Setengah Penganggur Pendidikan Tinggi

*Halaman ini sengaja dikosongkan*

Lampiran 4. Data Indikator Pasar Tenaga Kerja Sakernas Februari 2012 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	65.85	60.67	33.66	4.45	35.77	6.88	19.24	61.89	50.24	10.96	38.80	20.94
2	SUMATERA UTARA	74.55	69.85	30.67	3.34	35.79	6.13	24.06	65.98	51.13	11.15	37.70	20.58
3	SUMATERA BARAT	69.92	65.55	28.42	4.07	40.15	9.93	17.44	67.51	42.35	13.79	43.87	20.71
4	RIAU	66.91	63.46	37.80	5.85	34.90	6.31	15.15	56.36	44.80	11.23	43.98	24.93
5	JAMBI	69.40	66.86	37.12	4.42	33.60	6.75	18.12	58.47	56.24	8.01	35.76	26.09
6	SUMATERA SELATAN	73.55	69.44	28.99	2.78	39.19	3.79	25.25	68.23	55.21	10.46	34.32	21.64
7	BENGKULU	74.85	73.25	24.14	2.86	42.94	6.32	23.73	72.99	58.53	8.62	32.85	15.85
8	LAMPUNG	71.98	68.30	22.89	3.20	39.45	10.95	23.51	73.91	51.84	12.22	35.94	21.91
9	KEP. BANGKA BELITUNG	68.11	66.22	43.37	7.83	33.51	2.53	12.77	48.81	29.45	32.08	38.46	16.98
10	KEP. RIAU	69.33	65.26	62.86	2.86	24.33	1.93	8.02	34.28	15.06	23.98	60.95	12.88
11	DKI JAKARTA	70.83	63.23	64.68	4.46	23.11	1.59	6.15	30.86	2.21	17.52	80.27	6.79
12	JAWA BARAT	64.26	57.98	39.71	3.51	32.40	15.12	9.27	56.78	19.96	29.42	50.62	11.88
13	JAWA TENGAH	71.58	67.38	28.97	3.33	37.17	14.44	16.10	67.70	32.65	26.04	41.31	17.60
14	DI YOGYAKARTA	70.47	67.59	38.61	3.96	34.31	7.38	15.73	57.42	24.24	21.74	54.02	18.09
15	JAWA TIMUR	69.55	66.68	28.90	3.44	35.01	13.35	19.29	67.65	40.41	19.87	39.71	21.63
16	BANTEN	69.36	61.91	49.90	4.13	25.97	12.16	7.83	45.96	15.20	27.65	57.15	9.83
17	BALI	77.42	75.79	36.36	3.24	33.43	8.38	18.58	60.39	28.94	21.77	49.28	28.54
18	NUSA TENGGARA BARAT	69.27	65.66	21.63	1.79	43.10	13.22	20.25	76.58	45.34	13.33	41.33	17.63
19	NUSA TENGGARA TIMUR	74.77	72.99	16.98	1.15	41.39	3.68	36.81	81.87	68.16	7.81	24.03	29.51
20	KALIMANTAN BARAT	74.50	72.00	26.17	2.90	40.81	2.34	27.78	70.93	63.59	10.96	25.45	24.75
21	KALIMANTAN TENGAH	73.79	71.79	36.65	3.03	35.67	2.43	22.21	60.31	53.46	13.93	32.60	16.19
22	KALIMANTAN SELATAN	71.24	68.16	33.35	3.45	38.58	5.79	18.82	63.20	38.20	21.75	40.05	22.93
23	KALIMANTAN TIMUR	69.89	63.40	50.09	3.91	30.23	2.52	13.24	45.99	28.95	20.64	50.41	14.21
24	SULAWESI UTARA	66.82	61.27	34.18	3.82	39.87	10.25	11.88	62.00	33.98	17.35	48.67	16.67
25	SULAWESI TENGAH	74.63	71.84	26.29	3.90	39.31	8.29	22.21	69.82	48.91	14.74	36.35	22.45
26	SULAWESI SELATAN	64.56	60.39	28.96	4.10	42.68	3.89	20.38	66.94	43.12	14.10	42.77	23.75
27	SULAWESI TENGGARA	73.10	70.84	28.59	2.96	39.48	3.49	25.48	68.45	44.98	12.54	42.47	20.50
28	GORONTALO	64.36	61.27	31.85	3.57	40.31	12.08	12.19	64.58	36.52	22.31	41.18	20.84
29	SULAWESI BARAT	72.64	71.13	22.32	2.73	43.91	4.84	26.21	74.96	57.45	9.95	32.59	32.25
30	MALUKU	66.98	62.21	25.69	1.66	44.51	3.59	24.55	72.64	52.25	10.67	37.08	20.35
31	MALUKU UTARA	67.82	64.22	26.76	2.99	41.64	4.23	24.37	70.25	55.04	10.23	34.73	17.79
32	PAPUA BARAT	72.27	67.52	35.01	2.59	37.16	1.78	23.47	62.41	47.63	12.21	40.16	17.47
33	PAPUA	79.27	76.97	18.12	1.53	43.65	0.91	35.80	80.35	72.78	4.49	22.74	17.68
	INDONESIA	69.66	65.25	33.81	3.48	35.38	10.04	17.29	62.71	36.53	19.71	43.75	18.33



Lampiran 4 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	7.86	15.92	20.03	13.73	22.09	58.76	7.88	5.65	13.61	4.88	61.68	26.91	11.40
2	SUMATERA UTARA	5.40	12.97	17.17	13.86	27.17	60.69	6.31	3.73	11.32	7.04	66.09	26.86	7.05
3	SUMATERA BARAT	8.25	13.62	13.58	11.20	24.93	59.54	6.25	4.43	10.32	5.79	62.52	26.12	11.36
4	RIAU	9.12	14.35	16.78	8.58	24.26	50.38	5.17	4.16	8.29	0.91	68.42	24.48	7.10
5	JAMBI	6.24	15.53	21.34	17.27	22.32	54.35	3.65	1.67	8.19	6.20	70.86	20.29	8.84
6	SUMATERA SELATAN	5.44	12.19	18.98	14.47	23.73	64.03	5.59	3.65	10.44	7.12	66.30	25.45	8.25
7	BENGKULU	5.29	11.33	13.00	13.04	29.83	66.96	2.14	1.42	3.26	4.94	67.79	24.32	7.89
8	LAMPUNG	8.19	12.85	14.35	10.90	24.52	67.26	5.12	2.97	11.96	6.09	80.34	15.03	4.63
9	KEP. BANGKA BELITUNG	3.47	9.50	16.64	10.63	34.38	42.36	2.78	1.71	5.50	2.67	71.54	22.34	6.12
10	KEP. RIAU	4.23	8.21	10.52	7.92	34.95	31.33	5.87	7.70	5.12	1.34	69.61	20.85	9.54
11	DKI JAKARTA	2.25	4.09	5.52	4.62	48.98	24.14	10.72	7.83	13.55	9.97	46.73	31.51	21.76
12	JAWA BARAT	3.66	8.38	11.05	8.48	35.38	48.64	9.78	8.23	13.62	10.47	82.44	12.67	4.90
13	JAWA TENGAH	6.02	9.62	13.34	10.48	29.65	59.54	5.88	5.16	8.35	6.95	80.18	14.06	5.76
14	DI YOGYAKARTA	7.92	7.10	11.53	10.18	32.09	47.65	4.09	2.94	5.46	5.06	50.44	39.84	9.71
15	JAWA TIMUR	7.01	13.11	15.16	9.35	25.56	60.06	4.13	3.22	6.80	5.37	49.82	15.26	4.93
16	BANTEN	3.80	7.67	8.66	6.88	43.64	39.78	10.74	10.23	13.61	5.44	85.01	11.56	3.42
17	BALI	8.61	13.50	18.84	8.78	23.79	50.61	2.11	0.42	4.90	3.86	62.89	29.47	7.64
18	NUSA TENGGARA BARAT	9.92	13.57	15.74	9.97	21.64	69.80	5.21	4.94	6.24	4.86	71.92	20.25	7.83
19	NUSA TENGGARA TIMUR	8.08	17.02	21.34	13.37	24.86	78.44	2.39	1.23	6.12	7.17	78.96	15.60	5.44
20	KALIMANTAN BARAT	6.01	15.10	19.59	11.80	25.21	66.80	3.36	2.03	6.59	11.48	80.77	16.08	3.14
21	KALIMANTAN TENGAH	5.77	11.19	12.71	11.31	29.67	55.69	2.71	2.01	5.28	2.61	69.88	18.38	11.74
22	KALIMANTAN SELATAN	5.35	15.52	17.55	11.84	25.63	56.01	4.32	3.10	7.39	7.17	75.97	17.07	6.96
23	KALIMANTAN TIMUR	4.38	8.60	10.63	12.78	26.95	40.34	9.29	8.53	11.41	6.52	66.36	21.90	11.74
24	SULAWESI UTARA	5.88	11.65	13.24	11.38	31.15	56.21	8.32	6.22	11.40	11.07	64.85	31.68	3.47
25	SULAWESI TENGAH	9.68	12.48	16.98	15.01	22.62	63.87	3.73	2.26	7.62	4.90	68.73	24.46	6.81
26	SULAWESI SELATAN	10.48	13.52	15.19	12.11	21.27	59.94	6.46	5.00	10.25	7.17	68.71	19.82	11.48
27	SULAWESI TENGGARA	9.69	14.96	15.30	13.49	21.85	61.59	3.10	1.47	5.09	7.02	65.66	24.92	9.41
28	GORONTALO	7.09	11.09	14.20	11.19	24.53	59.14	4.81	4.35	7.45	3.39	82.18	14.85	2.96
29	SULAWESI BARAT	15.04	17.95	19.84	10.01	21.26	67.84	2.07	1.11	7.11	0.00	77.19	17.17	5.64
30	MALUKU	5.06	12.45	16.67	16.20	24.35	66.94	7.11	2.19	14.78	14.48	65.28	29.12	5.60
31	MALUKU UTARA	7.33	13.72	17.98	16.73	26.09	66.42	5.31	3.23	8.96	9.67	75.34	20.92	3.74
32	PAPUA BARAT	4.69	12.02	15.72	14.31	29.51	57.65	6.57	3.02	11.68	12.20	76.01	18.11	5.88
33	PAPUA	4.38	12.09	21.59	20.03	28.84	78.04	2.90	1.49	7.94	5.55	85.13	12.77	2.10
INDONESIA		6.08	11.23	14.21	10.45	29.40	55.84	6.32	4.83	10.05	7.12	74.86	18.55	6.59

Lampiran 5 Data Indikator Pasar Tenaga Kerja Sakernas Februari 2013 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	65.56	60.06	32.67	4.83	36.85	7.63	18.01	62.50	50.83	10.65	38.52	26.54
2	SUMATERA UTARA	72.72	68.35	32.60	3.58	34.37	6.27	23.18	63.82	49.39	11.72	38.89	20.87
3	SUMATERA BARAT	70.32	65.88	29.81	5.61	38.02	10.50	16.05	64.57	41.23	14.64	44.14	20.76
4	RIAU	69.51	66.64	41.61	4.29	32.15	7.33	14.61	54.10	46.28	11.48	42.25	30.56
5	JAMBI	69.21	67.20	35.34	4.60	33.14	8.13	18.79	60.05	54.19	9.11	36.70	39.27
6	SUMATERA SELATAN	71.94	67.99	30.83	3.17	37.55	5.03	23.42	66.00	52.96	11.50	35.53	24.88
7	BENGKULU	74.33	72.76	25.35	3.60	38.86	5.92	26.28	71.05	60.22	8.46	31.32	24.44
8	LAMPUNG	70.44	66.86	25.03	3.51	38.98	11.57	20.92	71.46	49.61	12.14	38.25	24.27
9	KEP. BANGKA BELITUNG	71.08	68.74	41.74	8.70	31.84	4.13	13.59	49.56	26.18	29.85	43.96	22.95
10	KEP. RIAU	70.53	66.02	65.16	3.22	23.55	2.36	5.71	31.62	13.18	27.39	59.43	15.13
11	DKI JAKARTA	68.44	61.64	69.99	3.61	19.31	2.77	4.32	26.40	1.07	17.85	81.08	5.20
12	JAWA BARAT	64.01	58.31	43.99	3.43	29.91	14.39	8.28	52.58	19.61	28.97	51.43	12.63
13	JAWA TENGAH	70.61	66.68	33.34	3.44	34.31	15.15	13.76	63.22	30.30	28.49	41.21	16.83
14	DI YOGYAKARTA	69.27	66.64	40.05	4.05	33.33	8.99	13.59	55.90	23.43	21.18	55.39	19.65
15	JAWA TIMUR	70.12	67.32	30.70	3.63	34.49	12.30	18.87	65.67	38.25	21.96	39.79	19.81
16	BANTEN	68.62	61.70	53.60	3.04	25.54	9.68	8.13	43.36	14.35	29.00	56.65	11.13
17	BALI	78.91	77.42	38.74	3.22	32.48	9.68	15.87	58.04	24.69	22.68	52.63	20.48
18	NUSA TENGGARA BARAT	70.35	66.57	21.33	2.93	41.67	13.33	20.75	75.75	43.37	14.63	42.00	19.89
19	NUSA TENGGARA TIMUR	74.95	73.45	15.98	1.45	41.71	4.14	36.72	82.57	67.69	8.18	24.13	36.90
20	KALIMANTAN BARAT	72.91	70.66	28.85	3.18	38.10	2.91	26.96	67.97	60.60	11.59	28.11	29.92
21	KALIMANTAN TENGAH	72.63	71.31	35.01	2.86	36.76	4.25	21.12	62.13	53.80	13.86	32.34	21.28
22	KALIMANTAN SELATAN	71.88	69.07	34.11	3.08	37.62	5.75	19.44	62.81	38.60	22.55	38.86	25.67
23	KALIMANTAN TIMUR	69.60	63.43	56.35	4.21	25.53	2.13	11.78	39.44	26.66	21.65	51.69	11.84
24	SULAWESI UTARA	64.63	59.98	35.61	5.08	38.03	10.08	11.20	59.31	31.41	18.21	50.37	16.91
25	SULAWESI TENGAH	71.79	69.89	28.14	3.25	40.72	8.72	19.17	68.61	47.40	15.58	37.02	27.21
26	SULAWESI SELATAN	63.58	59.87	30.92	4.24	41.80	4.35	18.69	64.84	41.39	13.38	45.24	25.86
27	SULAWESI TENGGARA	69.58	67.17	29.90	3.18	39.05	3.24	24.62	66.92	42.26	15.27	42.47	27.82
28	GORONTALO	64.33	61.56	33.26	4.47	41.12	9.23	11.92	62.26	35.13	16.28	48.59	18.38
29	SULAWESI BARAT	72.41	70.96	22.54	2.53	43.92	4.39	26.61	74.92	58.77	9.30	31.93	40.08
30	MALUKU	68.09	63.51	25.01	2.23	44.52	1.57	26.67	72.76	50.09	10.71	39.20	29.74
31	MALUKU UTARA	67.90	64.16	32.59	2.72	41.43	4.36	18.90	64.69	48.77	11.58	39.65	29.09
32	PAPUA BARAT	68.25	65.20	35.91	1.73	35.57	2.49	24.29	62.36	48.42	11.07	40.51	20.92
33	PAPUA	80.25	78.00	16.95	1.27	42.39	1.78	37.61	81.77	73.05	4.49	22.46	27.00
	INDONESIA	69.21	65.12	36.45	3.53	33.78	10.02	16.22	60.02	35.05	20.59	44.36	19.43

Lampiran 5 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	8.07	18.10	19.35	11.86	20.87	58.36	8.38	6.15	11.48	11.23	59.19	30.94	9.87
2	SUMATERA UTARA	6.67	12.76	16.19	12.88	28.98	58.68	6.01	4.51	8.59	7.01	63.82	30.73	5.45
3	SUMATERA BARAT	8.40	13.97	14.31	10.20	24.97	57.82	6.33	5.11	9.37	5.17	67.26	22.66	10.07
4	RIAU	8.62	15.42	15.45	10.99	22.34	50.05	4.13	2.29	7.34	5.14	68.77	20.23	11.00
5	JAMBI	9.87	17.77	23.91	12.30	19.13	55.45	2.90	2.05	3.09	7.58	65.90	25.47	8.63
6	SUMATERA SELATAN	6.32	14.86	17.78	13.61	24.24	61.04	5.49	3.69	10.48	4.39	72.87	21.22	5.91
7	BENGKULU	6.87	12.95	13.90	11.17	30.69	65.87	2.12	1.37	3.47	3.67	66.48	26.76	6.76
8	LAMPUNG	7.61	14.26	15.67	10.27	24.16	64.56	5.09	3.88	9.12	3.78	75.96	18.61	5.43
9	KEP. BANGKA BELITUNG	4.37	10.37	13.96	9.51	34.59	43.67	3.30	2.42	3.40	10.19	71.01	27.10	1.89
10	KEP. RIAU	4.97	5.87	7.94	7.76	40.72	28.26	6.39	6.48	7.09	4.33	66.33	32.93	0.74
11	DKI JAKARTA	2.06	3.98	5.24	4.33	47.12	20.94	9.94	9.59	12.51	5.18	47.40	33.65	18.95
12	JAWA BARAT	4.25	8.58	10.74	9.28	35.38	45.71	8.90	8.02	12.36	5.52	81.52	13.50	4.97
13	JAWA TENGAH	5.93	9.89	12.52	9.32	30.28	55.44	5.57	5.12	7.78	4.05	82.43	12.85	4.72
14	DI YOGYAKARTA	6.54	8.15	9.56	9.70	36.86	46.32	3.80	1.37	3.53	11.18	58.85	33.62	7.52
15	JAWA TIMUR	6.88	12.04	14.22	9.77	26.98	58.24	4.00	3.10	6.59	4.63	77.86	17.19	4.95
16	BANTEN	3.25	8.00	10.49	7.14	43.97	37.11	10.10	10.50	11.62	4.72	78.10	16.53	5.37
17	BALI	4.92	8.44	10.64	8.55	27.29	48.54	1.89	1.94	2.24	0.89	66.97	20.85	12.18
18	NUSA TENGGARA BARAT	8.36	13.88	16.75	10.08	23.31	67.95	5.37	5.00	5.67	7.74	74.56	18.41	7.03
19	NUSA TENGGARA TIMUR	9.10	17.14	22.93	14.13	22.26	79.59	2.01	1.06	5.54	5.61	77.56	15.57	6.87
20	KALIMANTAN BARAT	4.61	16.44	17.92	13.43	28.34	63.65	3.09	2.38	5.18	5.22	76.37	16.62	7.01
21	KALIMANTAN TENGAH	6.45	12.16	14.46	10.49	28.21	56.58	1.82	1.83	1.72	1.97	72.08	21.75	6.17
22	KALIMANTAN SELATAN	6.44	15.50	16.06	11.73	24.80	55.88	3.91	3.80	5.03	1.86	81.07	14.45	4.48
23	KALIMANTAN TIMUR	3.35	6.69	9.85	9.74	31.41	34.25	8.87	10.34	8.22	5.20	59.38	31.98	8.64
24	SULAWESI UTARA	4.72	11.57	11.94	10.99	36.29	54.28	7.19	4.69	11.84	6.98	65.50	28.49	6.01
25	SULAWESI TENGAH	9.38	12.31	16.25	13.93	22.31	62.44	2.65	2.28	3.73	2.74	73.44	20.15	6.41
26	SULAWESI SELATAN	10.99	15.62	16.68	10.43	20.83	56.98	5.83	5.31	8.88	3.10	67.36	22.18	10.47
27	SULAWESI TENGGARA	8.22	15.49	14.93	11.46	22.67	59.93	3.47	1.87	5.59	6.80	70.15	21.08	8.77
28	GORONTALO	5.98	9.65	12.33	13.06	25.21	54.80	4.31	3.61	7.71	2.81	84.09	10.24	5.67
29	SULAWESI BARAT	19.40	16.23	18.62	8.83	18.02	67.50	2.00	1.11	4.86	3.62	76.36	18.97	4.67
30	MALUKU	7.64	14.72	17.60	15.16	23.41	67.04	6.73	3.35	13.51	7.67	58.40	34.30	7.30
31	MALUKU UTARA	6.95	15.38	21.32	13.79	20.09	60.24	5.51	2.45	10.77	7.66	61.77	30.01	8.23
32	PAPUA BARAT	3.94	13.11	14.83	12.33	29.97	58.42	4.47	2.81	7.24	5.18	64.56	31.16	4.28
33	PAPUA	7.03	16.02	22.27	18.96	22.72	79.47	2.81	0.98	8.56	7.78	83.65	14.44	1.91
INDONESIA		6.18	11.42	13.72	10.17	29.84	53.48	5.92	4.91	8.78	5.22	74.74	19.04	6.22

Lampiran 6 Data Indikator Pasar Tenaga Kerja Sakernas Februari 2014 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	65.32	60.91	32.51	3.91	37.27	7.29	19.03	63.59	47.15	13.39	39.46	28.95
2	SUMATERA UTARA	73.04	68.69	30.51	3.26	36.73	6.50	23.00	66.23	48.05	11.23	40.72	23.62
3	SUMATERA BARAT	70.58	66.12	30.60	5.20	40.12	8.94	15.14	64.20	41.68	14.24	44.08	24.33
4	RIAU	66.88	63.55	41.84	5.20	32.61	5.72	14.62	52.96	42.41	13.08	44.50	26.83
5	JAMBI	66.51	64.84	35.38	4.91	37.86	5.11	16.75	59.72	49.35	8.09	42.56	34.37
6	SUMATERA SELATAN	71.96	69.20	33.59	2.34	36.37	5.45	22.26	64.08	52.26	11.93	35.80	28.24
7	BENGKULU	74.38	73.17	26.21	3.11	39.78	6.48	24.42	70.67	58.30	7.71	33.99	26.37
8	LAMPUNG	70.55	66.96	25.18	3.54	38.29	11.19	21.79	71.28	48.51	12.98	38.51	29.13
9	KEP. BANGKA BELITUNG	66.84	65.05	40.94	7.44	35.46	4.30	11.85	51.61	28.54	30.73	40.73	21.82
10	KEP. RIAU	67.83	64.26	60.46	2.23	28.31	3.56	5.44	37.31	13.96	31.38	54.65	13.14
11	DKI JAKARTA	68.49	61.75	67.85	3.80	20.41	2.91	5.02	28.35	2.17	19.48	78.35	9.64
12	JAWA BARAT	64.36	58.79	43.86	3.30	29.20	14.78	8.87	52.85	19.80	29.68	50.52	15.62
13	JAWA TENGAH	70.93	67.06	34.25	3.69	34.29	13.69	14.09	62.06	30.99	28.58	40.44	21.32
14	DI YOGYAKARTA	71.84	70.28	41.81	4.10	32.11	5.13	16.85	54.09	25.42	20.04	54.54	23.75
15	JAWA TIMUR	70.52	67.69	30.74	3.54	34.20	12.71	18.80	65.71	36.86	21.48	41.65	23.98
16	BANTEN	66.47	59.91	56.45	4.42	24.57	6.38	8.17	39.13	14.42	28.31	57.27	12.83
17	BALI	78.61	77.53	38.27	3.64	33.31	8.67	16.10	58.09	24.82	24.66	50.52	20.30
18	NUSA TENGGARA BARAT	70.71	66.96	22.10	2.59	39.46	12.91	22.94	75.31	45.48	13.07	41.46	21.79
19	NUSA TENGGARA TIMUR	74.04	72.58	17.33	1.33	41.75	5.17	34.42	81.35	65.04	9.81	25.15	36.01
20	KALIMANTAN BARAT	72.21	70.38	30.93	3.50	37.89	4.46	23.21	35.56	57.19	12.28	30.54	30.48
21	KALIMANTAN TENGAH	72.93	70.95	38.66	2.22	37.47	3.86	17.80	59.13	51.59	13.99	34.42	24.45
22	KALIMANTAN SELATAN	72.95	70.01	33.55	3.02	41.66	5.27	16.50	63.43	36.84	22.70	40.46	27.89
23	KALIMANTAN TIMUR	69.23	63.08	54.32	3.14	27.56	2.71	12.28	42.54	24.27	21.38	54.36	17.01
24	SULAWESI UTARA	66.14	61.33	35.52	4.01	36.92	12.18	11.37	60.47	31.87	17.63	50.50	19.80
25	SULAWESI TENGAH	71.79	69.70	29.59	3.35	38.27	7.90	20.88	67.05	46.35	14.12	39.52	26.54
26	SULAWESI SELATAN	62.02	58.43	32.61	4.37	41.23	3.84	17.96	63.02	40.65	12.53	46.82	32.19
27	SULAWESI TENGGARA	71.05	69.54	28.11	3.86	40.83	3.00	24.19	68.02	41.90	13.56	44.54	27.31
28	GORONTALO	66.25	64.63	34.55	2.99	40.85	9.68	11.92	62.46	33.34	15.53	51.13	20.18
29	SULAWESI BARAT	71.18	70.04	27.75	2.66	39.06	5.77	24.76	69.59	59.95	9.92	30.13	40.53
30	MALUKU	66.84	62.44	29.08	1.87	44.97	1.73	22.35	69.05	50.55	10.44	39.01	24.11
31	MALUKU UTARA	66.43	62.67	32.02	1.95	43.55	5.10	17.39	66.03	47.83	11.99	40.18	25.57
32	PAPUA BARAT	71.05	68.42	36.61	1.63	34.95	2.17	24.64	61.77	48.83	12.04	39.13	25.72
33	PAPUA	80.54	77.73	16.69	1.28	41.74	2.33	37.96	82.03	73.43	3.84	22.72	30.24
	INDONESIA	69.17	65.23	36.68	3.51	33.90	9.72	16.19	59.81	34.56	20.76	44.68	22.34

Lampiran 6 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	9.93	15.92	17.75	13.14	21.78	59.28	6.75	4.70	10.09	7.47	50.39	36.03	13.58
2	SUMATERA UTARA	6.10	13.10	14.24	13.77	27.19	61.17	5.95	4.86	8.47	4.01	57.64	34.86	7.50
3	SUMATERA BARAT	7.02	13.87	12.71	11.76	26.27	58.01	6.32	4.45	10.57	5.66	65.12	22.48	12.40
4	RIAU	7.02	12.94	14.79	9.96	26.95	48.71	4.99	2.39	8.33	8.32	66.57	27.30	6.13
5	JAMBI	6.11	16.70	22.29	14.32	20.35	55.16	2.50	1.30	5.15	2.58	56.48	33.15	10.37
6	SUMATERA SELATAN	6.28	14.55	18.05	11.60	22.67	60.31	3.84	2.31	6.72	7.45	67.90	25.44	6.66
7	BENGKULU	7.15	11.03	15.94	11.58	28.33	64.95	1.62	0.75	2.43	4.84	64.77	27.79	7.44
8	LAMPUNG	7.13	16.00	14.95	10.82	24.85	64.10	5.08	3.80	9.37	3.46	70.27	24.13	5.59
9	KEP. BANGKA BELITUNG	3.38	10.85	13.88	9.90	35.90	46.36	2.67	1.44	5.27	4.96	75.72	19.73	4.56
10	KEP. RIAU	2.59	4.96	9.51	9.24	39.61	32.67	5.26	3.90	8.19	3.12	56.59	33.45	9.96
11	DKI JAKARTA	2.61	3.35	5.09	4.39	51.28	23.69	9.84	11.20	11.50	4.75	37.68	50.45	11.87
12	JAWA BARAT	4.43	9.59	11.29	8.16	35.86	45.98	8.66	8.20	10.80	5.51	79.31	13.82	6.87
13	JAWA TENGAH	6.34	10.56	12.05	9.17	31.64	55.11	5.45	4.82	7.59	5.75	49.41	15.10	5.49
14	DI YOGYAKARTA	7.36	10.24	11.30	10.03	31.96	44.91	2.16	0.85	3.29	3.42	53.90	28.20	17.90
15	JAWA TIMUR	7.25	11.88	14.22	10.12	27.90	58.37	4.02	3.12	7.57	2.22	79.53	15.47	5.00
16	BANTEN	3.06	7.14	8.83	6.05	46.91	33.92	9.87	12.30	9.01	1.85	72.01	19.76	8.23
17	BALI	4.30	8.17	10.49	9.49	27.98	49.71	1.37	0.21	3.24	1.86	54.76	37.74	7.50
18	NUSA TENGGARA BARAT	11.12	13.53	14.97	11.04	20.54	69.07	5.30	5.10	6.83	2.79	70.32	20.69	8.99
19	NUSA TENGGARA TIMUR	9.51	18.35	20.68	14.76	22.64	78.05	1.97	1.11	4.91	5.03	79.95	16.73	3.32
20	KALIMANTAN BARAT	3.94	19.04	15.49	10.99	29.29	61.53	2.53	1.38	6.67	3.22	74.69	17.69	7.61
21	KALIMANTAN TENGAH	5.90	13.11	15.30	10.83	28.91	53.33	2.71	2.04	3.00	6.48	64.63	26.61	8.76
22	KALIMANTAN SELATAN	5.55	13.70	17.84	13.73	24.46	56.51	4.03	3.21	7.76	2.29	80.53	15.56	3.91
23	KALIMANTAN TIMUR	3.92	7.54	10.90	11.14	27.31	36.16	8.89	8.43	10.00	7.39	53.79	37.28	8.92
24	SULAWESI UTARA	5.65	11.30	11.26	10.82	36.19	55.51	7.27	5.38	10.26	8.31	65.99	31.44	2.57
25	SULAWESI TENGAH	7.58	14.50	14.60	13.15	23.63	59.83	2.92	1.37	6.63	3.87	69.73	19.17	11.10
26	SULAWESI SELATAN	9.74	15.62	16.18	10.11	22.45	56.28	5.79	4.11	8.94	7.24	51.33	32.13	16.54
27	SULAWESI TENGGARA	9.39	14.19	15.22	12.38	20.80	60.71	2.13	0.68	3.60	5.36	51.16	30.50	18.34
28	GORONTALO	7.42	9.18	13.34	9.11	19.81	55.66	2.44	1.53	5.06	4.04	76.67	18.91	4.41
29	SULAWESI BARAT	15.38	19.41	19.80	11.34	16.26	64.86	1.60	0.52	2.91	6.97	68.52	27.99	3.49
30	MALUKU	7.42	14.62	15.90	13.00	26.04	64.85	6.59	3.26	10.78	13.80	65.03	31.15	3.82
31	MALUKU UTARA	6.82	14.98	17.08	14.63	24.81	61.54	5.65	2.22	9.65	12.26	67.98	24.56	7.46
32	PAPUA BARAT	5.26	11.71	19.05	14.55	25.43	58.91	3.70	1.78	6.45	6.96	67.82	24.60	7.58
33	PAPUA	5.08	18.03	26.27	17.88	18.95	80.10	3.48	1.72	9.17	8.55	88.43	10.07	1.50
INDONESIA		6.16	11.67	13.45	10.09	30.38	53.59	5.70	4.75	8.42	4.72	72.31	20.58	7.11

Lampiran 7 Data Indikator Pasar Tenaga Kerja Sakernas Februari 2015 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	66.37	61.25	39.42	5.19	33.47	7.50	14.43	55.39	41.20	14.55	44.25	25.51
2	SUMATERA UTARA	69.90	65.43	37.43	2.67	34.12	8.65	17.12	59.90	40.24	15.63	44.13	22.96
3	SUMATERA BARAT	68.73	64.61	31.79	5.13	36.97	10.42	15.69	63.08	39.00	14.92	46.08	23.75
4	RIAU	68.85	64.22	44.15	4.19	30.13	8.47	13.06	51.67	46.09	11.18	42.72	27.24
5	JAMBI	69.92	68.01	40.26	4.24	34.07	6.02	15.41	55.50	49.88	13.60	36.53	31.71
6	SUMATERA SELATAN	70.54	66.99	33.98	2.54	37.10	6.07	20.30	63.47	49.83	13.24	36.93	24.29
7	BENGKULU	73.24	70.89	28.15	3.12	38.52	7.95	22.27	68.73	46.48	12.26	41.26	22.92
8	LAMPUNG	69.95	67.55	26.68	3.16	38.77	12.25	19.14	70.16	47.12	15.81	37.07	27.80
9	KEP. BANGKA BELITUNG	70.20	67.85	43.60	6.03	35.33	3.71	11.33	50.37	28.64	24.24	47.12	17.77
10	KEP. RIAU	66.16	60.18	68.48	2.69	20.87	3.60	4.36	28.83	14.56	32.03	53.42	12.85
11	DKI JAKARTA	72.60	66.53	68.03	4.67	19.10	2.60	5.60	27.30	0.77	19.66	79.58	10.88
12	JAWA BARAT	66.08	60.52	44.70	3.22	30.05	13.18	8.84	52.08	20.37	29.04	50.59	14.03
13	JAWA TENGAH	72.19	68.36	35.13	3.29	34.87	13.01	13.69	61.58	31.11	27.94	40.96	21.55
14	DI YOGYAKARTA	73.10	70.13	41.94	3.92	30.06	9.46	14.61	54.13	25.10	26.50	48.40	18.50
15	JAWA TIMUR	69.58	66.58	32.66	4.03	33.60	13.09	16.63	63.32	36.42	22.15	41.42	24.14
16	BANTEN	67.28	61.51	58.95	4.41	24.49	6.15	6.01	36.64	13.35	31.80	54.85	13.13
17	BALI	78.86	77.78	43.64	3.69	32.38	6.45	13.85	52.67	23.48	24.23	52.29	20.23
18	NUSA TENGGARA BARAT	71.66	68.09	20.56	2.58	43.52	14.79	18.55	76.86	44.56	16.33	39.11	18.48
19	NUSA TENGGARA TIMUR	72.95	70.67	20.42	1.88	40.15	3.79	33.76	77.70	63.30	7.53	29.17	32.73
20	KALIMANTAN BARAT	70.73	67.35	33.38	2.74	37.62	6.43	19.83	63.88	51.30	12.60	36.10	27.42
21	KALIMANTAN TENGAH	73.05	70.75	38.83	3.84	34.65	4.33	18.34	57.32	42.76	19.36	37.88	19.33
22	KALIMANTAN SELATAN	73.21	69.68	32.25	2.68	39.74	4.97	20.36	65.08	39.55	15.66	44.79	28.09
23	KALIMANTAN TIMUR	67.81	62.95	58.94	2.05	27.11	2.90	9.00	39.01	20.49	19.05	60.46	16.38
24	SULAWESI UTARA	66.24	60.48	34.22	4.42	38.80	11.55	11.02	61.36	34.48	12.65	52.87	20.48
25	SULAWESI TENGAH	70.21	68.11	26.73	4.07	39.70	9.44	20.06	69.20	50.68	10.97	38.35	27.66
26	SULAWESI SELATAN	62.23	58.61	33.07	3.49	40.92	4.62	17.90	63.44	40.97	13.61	45.42	26.23
27	SULAWESI TENGGARA	71.04	68.47	30.13	3.90	40.21	3.33	22.44	65.97	39.23	14.28	46.49	24.46
28	GORONTALO	66.37	64.34	30.36	3.73	40.60	9.18	16.12	65.90	33.79	14.54	51.66	20.97
29	SULAWESI BARAT	74.74	73.39	22.11	2.32	45.00	7.15	23.42	75.58	56.18	12.60	31.22	43.42
30	MALUKU	63.71	59.43	33.14	1.45	46.12	3.85	15.44	65.41	40.86	10.51	48.64	23.27
31	MALUKU UTARA	67.99	64.21	28.66	3.03	42.44	6.13	19.73	68.30	49.99	9.92	40.09	20.95
32	PAPUA BARAT	68.81	65.64	37.99	1.63	35.08	4.73	20.57	60.38	43.27	12.59	44.14	19.64
33	PAPUA	79.26	76.31	21.57	1.00	42.13	1.17	34.12	77.43	68.76	6.19	25.05	30.29
	INDONESIA	69.50	65.46	38.58	3.48	33.47	9.83	14.64	57.94	33.20	21.37	45.42	21.21

Lampiran 7 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	7.87	13.60	16.55	12.94	24.34	51.28	7.73	4.40	11.83	11.18	53.56	31.58	14.86
2	SUMATERA UTARA	5.73	11.76	14.70	12.28	27.07	54.58	6.39	4.80	8.06	9.19	52.48	37.73	9.79
3	SUMATERA BARAT	8.19	15.02	14.15	10.61	24.05	55.08	5.99	3.08	9.69	10.63	67.04	19.51	13.46
4	RIAU	8.01	15.28	13.44	12.46	23.51	47.81	6.72	3.30	10.82	12.03	65.02	24.04	10.94
5	JAMBI	5.97	18.32	19.06	13.23	23.17	51.55	2.73	1.59	5.60	1.98	57.56	31.57	10.86
6	SUMATERA SELATAN	5.69	13.81	15.03	12.41	24.87	58.75	5.03	2.59	10.72	6.04	67.01	26.45	6.55
7	BENGKULU	8.02	11.74	13.31	14.50	25.84	62.65	3.21	0.96	4.45	10.31	53.76	31.30	14.94
8	LAMPUNG	7.27	14.34	14.37	9.84	25.85	64.24	3.44	2.10	6.07	6.56	61.23	28.70	10.07
9	KEP. BANGKA BELITUNG	4.71	9.76	10.85	13.80	33.67	44.94	3.35	2.35	5.40	3.71	66.98	22.81	10.22
10	KEP. RIAU	4.57	5.40	6.43	8.20	46.66	25.93	9.05	4.87	12.94	8.92	61.83	26.31	11.86
11	DKI JAKARTA	4.48	4.53	4.99	5.51	48.03	21.67	8.36	10.46	8.69	4.51	34.14	52.34	13.52
12	JAWA BARAT	4.11	8.06	9.41	9.26	35.55	45.52	8.40	7.62	11.59	4.92	76.63	15.81	7.56
13	JAWA TENGAH	5.79	10.41	12.17	9.27	31.44	54.74	5.31	5.45	5.53	3.31	82.27	13.22	4.52
14	DI YOGYAKARTA	5.79	8.41	9.14	9.90	36.21	46.45	4.07	2.00	6.21	7.02	53.87	33.77	12.37
15	JAWA TIMUR	7.24	11.66	13.49	9.86	27.61	56.66	4.31	3.16	7.38	4.58	78.52	15.23	6.26
16	BANTEN	4.05	6.82	7.64	9.80	44.94	31.31	8.58	8.78	10.39	3.96	79.95	16.39	3.66
17	BALI	5.03	7.88	10.20	8.30	28.25	43.61	1.37	1.03	1.90	1.43	60.90	33.57	5.54
18	NUSA TENGGARA BARAT	9.85	13.25	14.80	11.85	19.95	70.29	4.98	3.71	7.97	8.18	74.81	17.31	7.87
19	NUSA TENGGARA TIMUR	8.51	15.66	18.71	14.53	25.64	74.83	3.12	1.35	6.65	8.78	73.59	22.31	4.10
20	KALIMANTAN BARAT	6.54	15.37	14.85	13.08	26.28	60.11	4.78	4.11	6.47	5.58	68.49	24.08	7.44
21	KALIMANTAN TENGAH	6.18	11.22	11.48	11.69	24.59	51.36	3.14	2.05	4.25	7.23	64.58	29.06	6.36
22	KALIMANTAN SELATAN	7.59	14.18	15.46	12.14	24.25	56.91	4.83	3.72	6.55	8.88	76.75	14.61	8.64
23	KALIMANTAN TIMUR	4.29	8.34	9.42	10.55	27.76	33.89	7.17	6.15	9.94	3.82	57.29	38.06	4.64
24	SULAWESI UTARA	5.42	10.19	14.08	12.54	33.67	56.36	8.69	4.96	14.06	10.20	57.34	34.53	8.13
25	SULAWESI TENGAH	9.24	15.44	15.36	12.56	21.06	63.02	2.99	2.02	4.53	5.36	71.47	22.50	6.03
26	SULAWESI SELATAN	10.59	13.03	12.74	8.89	25.87	56.70	5.81	3.61	10.90	5.98	58.98	27.94	13.09
27	SULAWESI TENGGARA	8.63	11.73	13.99	12.28	20.91	58.03	3.62	1.77	4.27	7.94	53.00	32.68	14.31
28	GORONTALO	8.41	9.36	10.68	9.62	20.76	57.58	3.06	1.60	7.31	4.02	75.74	16.93	7.32
29	SULAWESI BARAT	16.18	17.33	19.73	12.30	14.05	69.67	1.81	0.82	3.83	6.35	57.78	33.33	8.89
30	MALUKU	6.62	13.40	1.13	12.64	22.52	61.16	6.72	2.37	8.93	17.16	53.25	36.52	10.24
31	MALUKU UTARA	6.97	14.67	14.98	14.53	25.06	63.99	5.56	3.13	8.54	10.40	62.93	30.63	6.44
32	PAPUA BARAT	5.53	9.98	16.06	13.09	29.36	54.52	4.61	2.12	7.24	8.98	59.78	31.30	8.92
33	PAPUA	7.99	12.15	25.19	17.65	23.26	75.49	3.72	1.16	10.08	8.44	80.91	14.70	4.39
INDONESIA		6.24	10.89	12.39	10.35	30.34	51.85	5.81	4.64	8.50	5.86	70.38	21.72	7.90

Lampiran 8 Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2012 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	61.77	56.15	35.33	4.89	36.22	7.22	16.34	59.78	46.86	12.35	40.80	27.40
2	SUMATERA UTARA	69.41	65.11	36.49	3.61	32.64	8.24	19.02	59.91	43.40	15.04	41.57	19.51
3	SUMATERA BARAT	64.47	60.27	30.88	4.80	38.65	11.31	14.36	64.32	40.60	15.18	44.22	24.51
4	RIAU	62.90	60.20	42.48	5.19	32.74	5.48	14.11	52.33	44.73	13.41	41.87	23.51
5	JAMBI	65.07	62.97	36.03	4.46	35.08	7.38	17.06	59.51	55.04	9.85	35.12	32.02
6	SUMATERA SELATAN	69.56	65.60	33.28	2.68	36.11	5.40	22.53	64.04	56.37	11.16	32.48	28.38
7	BENGKULU	70.07	67.54	29.29	3.35	38.78	6.40	22.17	67.36	52.62	10.09	37.29	25.03
8	LAMPUNG	66.27	62.83	26.55	3.88	37.66	10.61	21.31	69.57	48.31	16.02	35.68	24.17
9	KEP. BANGKA BELITUNG	65.67	63.38	44.94	6.68	32.60	5.35	10.44	48.38	28.57	34.00	37.43	19.47
10	KEP. RIAU	66.25	62.69	65.37	4.76	22.83	1.94	5.10	29.87	11.93	33.58	54.50	9.06
11	DKI JAKARTA	71.56	64.50	68.29	3.87	20.54	2.12	5.18	27.84	0.53	18.69	80.80	5.87
12	JAWA BARAT	63.78	57.99	41.86	3.44	30.80	14.82	9.08	54.70	21.65	29.43	48.91	13.36
13	JAWA TENGAH	71.43	67.41	30.63	3.23	35.97	13.96	16.21	66.14	31.39	28.65	39.96	19.15
14	DI YOGYAKARTA	70.85	68.04	39.06	4.38	31.48	8.70	16.38	56.56	26.91	23.12	49.97	20.48
15	JAWA TIMUR	69.62	66.75	30.42	3.38	33.37	13.51	19.32	66.20	39.16	22.33	38.50	20.79
16	BANTEN	65.03	58.44	55.57	2.95	24.42	10.03	7.04	41.49	13.09	32.74	54.16	10.03
17	BALI	76.97	75.40	42.93	4.01	33.48	9.24	14.68	53.05	25.24	22.53	52.24	17.37
18	NUSA TENGGARA BARAT	66.02	62.55	21.12	2.74	42.04	18.64	17.05	76.13	44.25	15.91	39.84	18.49
19	NUSA TENGGARA TIMUR	70.58	68.55	18.56	1.56	42.30	4.47	30.56	79.88	61.61	12.97	25.41	31.41
20	KALIMANTAN BARAT	71.77	69.28	29.74	2.74	38.60	3.41	26.95	67.52	59.50	13.11	27.39	27.88
21	KALIMANTAN TENGAH	69.90	67.68	36.19	3.31	36.17	2.69	22.25	60.50	55.41	13.75	30.83	23.52
22	KALIMANTAN SELATAN	71.93	68.16	32.67	3.59	40.20	6.49	19.95	63.74	41.43	17.64	40.92	26.86
23	KALIMANTAN TIMUR	66.64	60.71	52.71	4.33	29.00	3.58	12.15	42.96	28.35	22.84	48.81	13.95
24	SULAWESI UTARA	61.93	57.11	39.06	3.99	40.00	10.87	10.07	56.95	32.61	17.52	49.88	18.22
25	SULAWESI TENGAH	66.38	63.78	28.94	4.61	36.33	8.71	18.70	66.44	49.88	13.65	36.47	25.91
26	SULAWESI SELATAN	62.82	59.14	31.79	4.28	40.73	5.28	19.96	63.92	44.03	13.35	42.62	25.28
27	SULAWESI TENGGARA	67.35	64.63	29.48	3.43	38.03	5.73	22.41	67.08	40.93	16.34	42.72	27.95
28	GORONTALO	63.08	60.32	31.50	5.14	41.01	11.30	14.09	63.35	37.80	19.53	42.66	23.29
29	SULAWESI BARAT	71.73	70.20	23.33	1.93	40.75	5.52	25.44	74.74	57.27	9.90	32.83	37.56
30	MALUKU	63.71	58.92	26.52	1.78	46.29	2.65	22.87	71.70	48.99	11.59	39.42	22.33
31	MALUKU UTARA	66.35	63.19	25.62	2.90	41.43	5.17	24.63	71.47	55.00	9.94	35.07	26.24
32	PAPUA BARAT	67.12	63.44	36.49	<b>2.28</b>	35.60	1.68	21.03	61.23	46.52	13.20	40.30	18.02
33	PAPUA	78.91	76.05	17.91	1.05	42.98	1.89	37.40	81.05	72.83	5.16	22.00	21.67
	INDONESIA	67.88	63.71	36.36	3.50	34.81	10.42	16.15	60.14	35.09	21.66	43.24	19.42



Lampiran 8 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	9.45	20.79	20.38	11.66	17.26	55.86	9.10	6.26	15.46	6.88	64.21	28.65	7.14
2	SUMATERA UTARA	4.99	11.61	13.91	12.15	32.64	54.12	6.20	4.68	8.78	6.73	62.54	32.15	5.30
3	SUMATERA BARAT	7.94	14.75	16.86	11.81	21.78	57.11	6.20	4.68	10.23	7.15	68.96	24.14	6.90
4	RIAU	7.03	15.31	14.50	9.18	26.13	47.78	4.30	3.18	6.44	4.59	72.31	24.00	3.70
5	JAMBI	6.61	18.12	22.22	14.27	18.69	55.67	3.22	1.82	6.80	4.23	74.40	20.36	5.24
6	SUMATERA SELATAN	5.85	17.62	18.94	12.66	22.68	60.55	5.70	3.79	10.31	8.01	76.26	19.45	4.30
7	BENGKULU	6.85	13.90	16.93	11.99	22.72	62.18	3.61	2.09	7.00	3.93	71.22	24.51	4.27
8	LAMPUNG	8.11	14.43	13.63	8.86	22.46	62.95	5.18	3.51	10.17	5.08	72.44	22.61	4.95
9	KEP. BANGKA BELITUNG	4.21	10.42	15.06	10.64	31.88	43.76	3.49	2.57	5.29	5.10	70.18	25.77	4.05
10	KEP. RIAU	2.61	4.93	5.69	6.30	37.88	25.90	5.37	3.75	7.38	3.09	66.49	27.53	5.98
11	DKI JAKARTA	1.75	3.37	4.18	6.73	49.63	21.57	9.87	9.17	10.33	10.21	47.26	37.43	15.31
12	JAWA BARAT	4.10	8.72	10.64	9.63	34.91	47.62	9.08	8.25	12.71	5.29	83.01	13.08	3.91
13	JAWA TENGAH	5.91	10.52	12.98	10.46	28.81	57.10	5.63	4.61	9.66	5.29	82.21	13.87	3.92
14	DI YOGYAKARTA	6.83	9.80	11.61	10.16	31.50	47.89	3.97	2.14	6.09	5.76	50.33	39.37	10.30
15	JAWA TIMUR	6.76	11.73	14.10	9.99	27.95	58.95	4.12	2.87	8.23	4.32	82.86	13.69	3.45
16	BANTEN	3.28	6.82	8.82	8.98	43.40	35.68	10.13	9.63	11.84	7.96	83.90	10.67	5.43
17	BALI	4.49	7.33	10.41	8.22	27.62	44.89	2.04	1.24	3.09	2.89	66.68	26.85	6.46
18	NUSA TENGGARA BARAT	11.40	15.94	17.15	10.45	18.53	68.66	5.26	3.59	11.44	5.12	79.29	15.39	5.33
19	NUSA TENGGARA TIMUR	10.29	16.70	20.53	14.41	21.95	76.25	2.89	1.80	7.32	4.86	83.03	14.41	2.56
20	KALIMANTAN BARAT	4.71	18.94	20.17	11.49	25.44	64.12	3.48	2.62	7.02	3.07	82.37	15.39	2.24
21	KALIMANTAN TENGAH	5.13	13.07	17.53	11.43	25.86	56.65	3.17	2.05	6.05	5.60	78.79	17.34	3.87
22	KALIMANTAN SELATAN	7.01	16.07	16.44	11.06	22.11	57.22	5.25	4.59	8.20	3.46	78.11	18.78	3.11
23	KALIMANTAN TIMUR	3.55	7.71	10.59	9.48	27.57	37.59	8.90	8.90	10.66	3.29	71.06	21.98	6.97
24	SULAWESI UTARA	5.09	11.42	11.95	10.19	35.93	52.92	7.79	5.55	11.96	6.31	64.55	29.33	6.12
25	SULAWESI TENGAH	9.80	14.77	14.98	11.71	20.44	61.56	3.93	2.27	7.39	6.57	79.45	16.40	4.15
26	SULAWESI SELATAN	10.62	14.64	13.67	9.46	21.35	58.05	5.87	4.40	9.37	6.45	71.29	22.30	6.41
27	SULAWESI TENGGARA	10.26	16.49	17.07	9.98	17.46	59.96	4.04	1.75	7.90	7.61	67.93	24.84	7.23
28	GORONTALO	6.80	11.88	15.05	9.34	17.99	58.41	4.36	3.31	7.94	5.92	87.84	11.20	0.95
29	SULAWESI BARAT	19.16	19.76	17.52	9.39	13.05	67.96	2.14	1.96	3.11	1.49	74.66	22.37	2.98
30	MALUKU	6.42	15.41	16.96	12.55	22.54	67.22	7.51	2.85	13.62	15.00	65.67	30.83	3.50
31	MALUKU UTARA	9.04	16.29	18.30	14.28	20.38	67.50	4.76	3.26	8.27	5.42	72.07	23.39	4.54
32	PAPUA BARAT	6.01	13.42	15.21	12.99	24.27	56.78	5.49	2.96	10.06	6.46	72.99	20.51	6.50
33	PAPUA	3.58	12.62	21.94	20.75	27.30	78.90	3.63	1.42	10.28	9.80	84.45	13.16	2.38
INDONESIA		5.97	11.54	13.43	10.36	29.43	53.57	6.14	4.80	9.70	6.00	77.27	18.18	4.55

Lampiran 9 Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2013 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	62.07	55.68	34.62	5.11	37.93	6.00	16.34	60.27	46.53	10.91	42.57	28.98
2	SUMATERA UTARA	70.67	66.05	36.45	3.44	31.23	7.60	21.28	60.11	43.45	14.67	41.88	20.23
3	SUMATERA BARAT	62.94	58.54	31.12	4.66	40.57	9.65	14.00	64.22	39.34	13.35	47.31	31.31
4	RIAU	63.62	60.12	41.34	4.83	32.98	6.47	14.38	53.83	44.26	14.05	41.69	28.02
5	JAMBI	62.66	59.63	37.13	4.35	38.21	6.70	13.61	58.52	52.37	10.16	37.47	40.74
6	SUMATERA SELATAN	66.50	63.17	33.30	2.70	37.21	5.08	21.70	63.99	54.69	10.43	34.88	37.38
7	BENGKULU	67.32	64.13	30.06	3.20	40.29	6.18	20.27	66.74	52.16	9.07	38.77	33.29
8	LAMPUNG	64.70	60.91	25.75	3.47	39.29	11.07	20.42	70.78	51.46	13.24	35.29	37.08
9	KEP. BANGKA BELITUNG	65.32	62.91	46.55	7.05	30.46	4.83	11.11	46.40	28.06	32.68	39.26	28.78
10	KEP. RIAU	65.58	61.48	68.35	4.10	20.40	3.00	4.16	27.55	10.15	37.70	52.15	12.65
11	DKI JAKARTA	68.09	61.95	67.29	3.67	20.16	2.44	6.43	29.04	0.33	18.77	80.90	7.34
12	JAWA BARAT	63.01	57.20	43.31	3.36	30.10	14.81	8.43	53.33	19.93	29.24	50.83	15.42
13	JAWA TENGAH	70.72	66.46	31.83	3.27	35.97	12.26	16.67	64.90	30.86	25.57	43.57	22.30
14	DI YOGYAKARTA	68.89	66.58	39.88	4.56	32.45	7.08	16.03	55.56	27.86	19.76	52.37	31.62
15	JAWA TIMUR	69.92	66.89	30.54	3.24	33.91	12.51	19.80	66.22	37.44	20.61	41.95	22.84
16	BANTEN	63.53	57.25	57.19	3.17	24.35	9.75	5.53	39.64	14.99	32.49	52.52	11.04
17	BALI	75.35	74.00	44.39	3.76	29.06	9.22	13.57	51.85	24.00	24.18	51.81	23.79
18	NUSA TENGGARA BARAT	65.44	61.92	21.89	2.62	39.51	19.23	16.75	75.49	45.02	15.23	39.75	21.72
19	NUSA TENGGARA TIMUR	68.72	66.54	19.15	1.55	46.19	3.11	29.99	79.29	60.90	12.11	26.99	40.81
20	KALIMANTAN BARAT	69.75	66.94	31.38	2.64	37.03	4.30	24.65	65.99	57.57	12.98	29.45	33.04
21	KALIMANTAN TENGAH	68.21	66.10	38.00	3.23	37.71	3.62	17.44	58.77	52.70	14.66	32.64	31.28
22	KALIMANTAN SELATAN	69.08	66.46	33.40	3.32	38.02	7.31	17.65	63.28	40.22	17.41	42.37	33.38
23	KALIMANTAN TIMUR	63.79	58.66	52.88	3.92	29.16	4.64	9.40	43.20	26.61	23.09	50.30	14.62
24	SULAWESI UTARA	59.76	55.77	39.94	3.58	35.03	12.51	8.94	56.47	34.23	15.93	49.84	20.97
25	SULAWESI TENGAH	65.92	63.11	29.06	3.81	40.58	7.19	19.37	67.13	49.25	12.48	38.27	33.94
26	SULAWESI SELATAN	60.49	57.40	33.57	4.11	38.10	5.86	18.35	62.32	43.39	12.71	43.90	34.55
27	SULAWESI TENGGARA	65.79	62.86	30.00	2.90	40.72	4.42	21.97	67.10	41.53	14.54	43.94	35.63
28	GORONTALO	62.00	59.44	32.33	3.35	42.47	9.31	12.55	64.33	36.66	16.59	46.75	26.15
29	SULAWESI BARAT	66.82	65.26	25.03	1.76	45.38	5.01	22.82	73.21	57.55	10.04	32.40	50.45
30	MALUKU	62.31	56.24	31.03	2.15	42.42	3.22	21.18	66.82	48.07	10.05	41.88	24.23
31	MALUKU UTARA	64.38	61.90	26.92	2.85	40.95	5.25	24.04	70.23	54.31	9.20	36.49	30.08
32	PAPUA BARAT	66.41	63.34	36.23	<b>2.02</b>	38.14	2.07	21.54	61.75	48.71	10.25	41.04	26.69
33	PAPUA	78.01	75.49	17.51	1.25	42.79	1.51	36.94	81.24	72.90	5.14	21.96	28.27
INDONESIA		66.90	62.72	37.03	3.39	33.73	9.95	15.90	59.58	34.36	20.60	45.04	23.39

Lampiran 9 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	10.35	19.57	17.65	11.31	18.76	56.55	10.30	9.06	16.21	10.29	61.35	29.12	9.53
2	SUMATERA UTARA	5.41	11.81	11.81	14.23	33.06	54.25	6.53	5.66	9.92	8.44	58.63	34.21	7.16
3	SUMATERA BARAT	11.14	17.96	14.36	8.42	18.34	56.61	6.99	6.92	11.71	7.91	64.76	26.02	9.21
4	RIAU	8.94	14.54	14.19	9.51	23.90	49.01	5.50	4.51	9.25	6.46	67.35	26.30	6.35
5	JAMBI	15.18	19.40	15.42	9.53	15.16	54.39	4.84	4.07	9.12	4.48	69.05	25.07	5.87
6	SUMATERA SELATAN	9.90	20.82	16.63	10.16	19.22	60.24	5.00	5.32	10.06	8.91	68.68	26.16	5.16
7	BENGKULU	10.88	16.60	15.45	8.99	20.01	61.66	4.74	4.25	8.04	5.72	66.73	27.25	6.02
8	LAMPUNG	13.20	18.48	15.17	7.43	18.86	65.01	5.85	5.10	11.30	7.52	70.91	24.50	4.59
9	KEP. BANGKA BELITUNG	8.27	16.31	13.20	8.28	27.33	41.08	3.70	3.22	7.25	3.71	72.02	24.18	3.80
10	KEP. RIAU	4.21	5.89	5.88	5.67	41.49	25.35	6.25	4.02	7.87	3.68	68.94	30.75	0.31
11	DKI JAKARTA	2.66	3.81	4.23	6.50	47.11	22.40	9.02	8.55	10.90	5.10	40.02	44.18	15.80
12	JAWA BARAT	5.49	11.14	9.52	8.87	32.10	45.15	9.22	8.79	12.61	5.52	83.73	12.42	3.85
13	JAWA TENGAH	8.39	12.93	10.29	11.94	25.40	55.29	6.02	6.08	10.80	4.42	79.88	16.43	3.69
14	DI YOGYAKARTA	13.32	13.46	13.04	8.67	24.16	47.36	3.34	3.69	5.96	4.19	54.80	37.87	7.34
15	JAWA TIMUR	7.81	12.89	12.39	11.71	27.51	57.54	4.33	4.48	8.77	3.53	80.44	16.22	3.34
16	BANTEN	4.17	7.38	8.50	7.11	43.17	34.52	9.90	10.03	12.74	5.62	77.00	16.53	6.47
17	BALI	5.74	9.49	11.87	8.40	28.43	44.40	1.79	2.15	2.89	2.72	58.22	33.80	7.98
18	NUSA TENGGARA BARAT	11.48	18.20	17.09	9.99	18.62	68.20	5.38	5.93	11.59	6.79	78.25	16.53	5.22
19	NUSA TENGGARA TIMUR	11.52	19.04	20.75	12.55	21.26	75.98	3.16	2.70	7.47	7.79	72.72	21.74	5.54
20	KALIMANTAN BARAT	7.96	20.84	15.02	9.97	24.06	62.85	4.03	4.17	9.00	4.73	78.90	16.94	4.15
21	KALIMANTAN TENGAH	7.85	16.46	16.36	10.12	23.10	54.61	3.09	2.80	6.57	6.06	71.18	23.58	5.23
22	KALIMANTAN SELATAN	9.22	16.44	16.53	10.94	21.48	56.92	3.79	3.91	7.39	2.67	75.06	18.58	6.36
23	KALIMANTAN TIMUR	3.01	7.97	9.69	9.70	26.77	38.39	8.04	7.38	10.37	5.78	60.90	31.99	7.11
24	SULAWESI UTARA	5.19	12.70	12.83	9.35	33.44	53.28	6.68	5.28	10.75	5.63	62.87	31.90	5.24
25	SULAWESI TENGAH	11.74	18.61	16.12	9.47	19.93	61.69	4.27	3.96	7.15	4.81	69.05	22.26	8.69
26	SULAWESI SELATAN	11.81	17.81	14.61	9.23	22.80	56.54	5.10	4.85	9.29	6.50	61.10	28.73	10.16
27	SULAWESI TENGGARA	14.27	18.81	16.08	7.56	16.50	60.43	4.46	3.63	8.53	5.53	58.41	30.39	11.21
28	GORONTALO	6.93	13.02	13.20	10.21	21.57	58.71	4.12	3.51	9.57	9.19	82.53	15.57	1.89
29	SULAWESI BARAT	23.95	25.74	14.45	6.48	10.27	67.40	2.33	2.35	4.81	2.65	68.23	24.73	7.04
30	MALUKU	7.70	14.24	15.73	14.40	20.94	63.49	9.75	8.55	17.16	14.75	62.41	32.60	4.99
31	MALUKU UTARA	8.50	16.41	17.55	13.01	22.43	66.91	3.86	2.59	6.19	7.12	66.24	25.94	7.82
32	PAPUA BARAT	5.12	14.16	19.43	12.29	22.52	56.92	4.62	4.37	8.23	6.42	61.98	23.38	14.64
33	PAPUA	5.75	14.48	26.14	17.15	23.67	78.72	3.23	3.01	8.43	8.99	89.78	8.64	1.57
INDONESIA		7.85	13.38	12.24	10.28	27.67	52.58	6.25	6.09	10.27	5.65	74.77	20.02	5.21

Lampiran 10 Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2014 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	63.06	57.38	35.35	4.50	37.60	6.89	15.67	60.15	44.09	12.27	43.64	29.49
2	SUMATERA UTARA	67.07	62.90	38.95	3.55	34.01	6.01	17.47	57.50	42.52	15.09	42.38	23.30
3	SUMATERA BARAT	65.19	60.95	31.11	4.38	41.64	9.93	12.94	64.51	37.55	14.32	48.13	25.15
4	RIAU	63.31	59.16	40.87	4.69	33.95	7.53	12.95	54.44	44.28	13.87	41.85	27.61
5	JAMBI	65.59	62.25	33.29	4.15	39.11	6.66	16.80	62.56	49.38	10.23	40.40	35.87
6	SUMATERA SELATAN	68.85	65.43	34.61	2.85	37.13	4.48	20.93	62.54	53.37	10.77	35.87	31.59
7	BENGKULU	68.29	65.92	30.91	3.30	39.61	6.37	19.80	65.79	50.62	9.39	39.99	25.95
8	LAMPUNG	66.99	63.78	26.22	3.11	39.36	11.87	19.45	70.67	48.87	13.47	37.66	29.91
9	KEP. BANGKA BELITUNG	65.45	62.09	43.21	5.75	31.63	6.92	12.49	51.04	31.41	28.74	39.85	21.04
10	KEP. RIAU	65.95	61.53	63.87	5.25	23.62	2.82	4.43	30.88	10.21	33.87	55.92	10.51
11	DKI JAKARTA	66.61	60.97	67.77	4.72	20.70	2.25	4.56	27.51	0.58	19.82	79.59	7.44
12	JAWA BARAT	62.77	57.47	42.45	3.54	31.73	14.18	8.10	54.01	19.87	29.07	51.06	16.88
13	JAWA TENGAH	69.68	65.72	31.72	3.86	36.53	13.18	14.71	64.42	31.26	27.54	41.20	22.44
14	DI YOGYAKARTA	71.05	68.69	43.22	3.90	30.51	7.62	14.75	52.88	25.41	22.32	52.27	20.35
15	JAWA TIMUR	68.12	65.27	30.99	3.93	35.61	13.03	16.45	65.08	37.61	21.83	40.56	24.90
16	BANTEN	63.84	58.05	57.26	3.18	23.53	9.92	6.12	39.57	12.46	33.83	53.71	12.40
17	BALI	74.91	73.48	43.35	4.33	30.07	8.94	13.31	52.32	23.26	23.75	53.00	20.04
18	NUSA TENGGARA BARAT	66.63	62.80	23.39	2.49	41.59	17.33	16.20	74.13	43.13	15.86	41.01	19.28
19	NUSA TENGGARA TIMUR	68.91	66.67	19.59	1.50	46.92	3.06	28.94	78.91	60.77	12.40	26.83	37.44
20	KALIMANTAN BARAT	69.93	67.10	31.26	3.32	35.37	4.97	25.09	65.42	57.76	13.04	29.21	29.40
21	KALIMANTAN TENGAH	68.56	66.34	37.13	3.51	38.61	3.48	17.27	59.36	53.11	13.06	33.83	28.89
22	KALIMANTAN SELATAN	69.46	66.82	34.04	2.93	39.61	6.32	17.11	63.04	39.81	15.91	44.28	30.59
23	KALIMANTAN TIMUR	64.10	59.37	52.14	3.79	29.95	3.99	10.12	44.07	27.84	22.78	49.38	17.69
24	SULAWESI UTARA	59.99	55.47	38.78	3.43	36.19	13.51	8.10	57.79	32.73	17.82	49.46	19.44
25	SULAWESI TENGAH	66.76	64.31	28.98	4.29	40.20	7.38	19.14	66.72	47.68	11.90	40.42	28.54
26	SULAWESI SELATAN	62.04	58.88	32.67	3.78	39.22	5.55	18.78	63.55	41.81	12.48	45.71	28.75
27	SULAWESI TENGGARA	66.87	63.91	28.36	3.78	39.84	5.42	22.59	67.86	42.62	13.83	43.55	29.80
28	GORONTALO	62.84	60.21	30.01	3.41	40.61	12.26	13.71	66.58	39.24	17.01	43.75	19.37
29	SULAWESI BARAT	71.06	69.58	24.81	2.01	40.99	6.59	25.59	73.18	56.84	10.69	32.47	41.15
30	MALUKU	60.92	54.52	32.45	1.15	44.63	3.08	18.69	66.40	48.09	8.79	43.12	24.10
31	MALUKU UTARA	63.88	60.50	26.95	3.36	43.37	5.37	20.95	69.69	52.51	9.48	38.02	27.22
32	PAPUA BARAT	68.30	64.87	35.65	<b>2.76</b>	37.07	3.49	21.03	61.59	45.28	12.37	42.35	22.77
33	PAPUA	78.67	75.96	16.50	0.82	41.93	1.60	39.15	82.67	70.59	4.59	24.82	27.73
	INDONESIA	66.60	62.64	36.97	3.65	34.69	10.03	14.66	59.38	34.00	21.16	44.84	22.76

Lampiran 10 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	8.62	18.35	17.65	11.45	21.10	55.79	9.02	5.91	14.07	9.95	56.71	31.53	11.75
2	SUMATERA UTARA	5.52	12.46	13.60	11.14	32.70	52.66	6.23	4.02	9.03	8.68	55.31	36.47	8.22
3	SUMATERA BARAT	7.57	14.35	14.99	10.26	24.32	57.15	6.50	4.64	9.89	7.69	65.86	23.07	11.07
4	RIAU	6.70	12.98	15.57	9.36	27.34	49.62	6.56	4.31	9.35	10.84	63.14	29.78	7.07
5	JAMBI	7.17	19.49	18.83	13.44	20.51	57.80	5.08	2.81	9.43	8.44	65.61	25.79	8.60
6	SUMATERA SELATAN	5.20	18.00	18.74	10.67	22.40	58.24	4.96	2.61	10.78	6.16	71.75	22.43	5.83
7	BENGKULU	6.43	13.70	15.93	12.06	26.04	61.17	3.47	1.90	6.15	5.31	62.80	28.71	8.50
8	LAMPUNG	6.54	16.42	14.39	10.56	24.98	65.27	4.79	2.95	9.66	5.85	70.79	22.02	7.19
9	KEP. BANGKA BELITUNG	4.54	11.82	12.59	11.03	33.30	45.51	5.14	3.60	8.44	6.66	70.14	24.00	5.85
10	KEP. RIAU	2.65	5.41	5.73	5.52	42.92	26.65	6.69	4.80	9.20	3.11	58.26	35.99	5.75
11	DKI JAKARTA	2.13	3.57	4.34	3.74	51.71	21.60	8.47	10.10	8.70	5.31	46.21	33.18	20.61
12	JAWA BARAT	3.76	9.59	11.64	8.41	36.17	46.80	8.45	6.55	14.27	4.78	82.94	12.96	4.10
13	JAWA TENGAH	6.07	11.03	12.50	9.08	30.34	56.78	5.68	4.74	9.67	3.42	80.51	15.93	3.56
14	DI YOGYAKARTA	7.46	8.36	9.61	9.33	34.87	46.06	3.33	1.42	5.65	4.48	51.02	38.21	10.78
15	JAWA TIMUR	6.84	12.47	14.27	9.53	28.00	58.37	4.19	2.71	8.73	3.74	80.22	15.44	4.35
16	BANTEN	4.12	7.47	8.83	7.15	45.07	34.36	9.07	7.98	12.35	4.85	75.74	18.08	6.18
17	BALI	4.31	7.88	10.39	8.87	30.62	44.39	1.90	0.90	3.09	2.89	58.75	32.98	8.27
18	NUSA TENGGARA BARAT	9.04	15.25	15.79	11.27	20.85	68.51	5.75	3.47	12.45	6.69	74.81	19.40	5.79
19	NUSA TENGGARA TIMUR	11.22	17.84	19.83	13.26	22.39	75.11	3.26	1.58	7.67	8.49	74.21	20.06	5.73
20	KALIMANTAN BARAT	5.14	18.04	16.19	11.38	28.59	62.26	4.04	2.58	8.56	5.78	80.39	15.35	4.26
21	KALIMANTAN TENGAH	5.56	15.34	16.61	11.70	24.26	55.24	3.24	2.09	6.22	4.37	69.47	22.69	7.84
22	KALIMANTAN SELATAN	6.82	14.32	16.92	11.67	24.78	56.93	3.80	2.54	7.60	4.11	76.63	17.42	5.96
23	KALIMANTAN TIMUR	3.81	8.70	11.11	10.63	27.54	38.79	7.38	5.21	10.95	5.87	66.12	26.77	7.11
24	SULAWESI UTARA	4.14	11.40	12.23	10.24	36.21	54.22	7.54	4.26	11.71	10.32	64.31	30.96	4.73
25	SULAWESI TENGAH	8.62	13.61	16.33	12.22	23.14	61.13	3.68	2.15	7.05	5.70	73.44	21.63	4.93
26	SULAWESI SELATAN	9.05	13.57	14.19	10.07	26.26	57.55	5.08	2.36	9.08	9.58	60.45	25.05	14.50
27	SULAWESI TENGGARA	9.37	15.34	16.54	11.08	21.17	61.07	4.43	1.77	8.94	7.01	59.35	30.52	10.13
28	GORONTALO	5.32	11.22	12.44	10.38	22.65	60.27	4.18	2.31	10.36	7.19	81.46	14.86	3.69
29	SULAWESI BARAT	15.45	19.91	17.10	10.69	16.12	67.64	2.08	1.93	2.71	1.99	58.00	32.82	9.18
30	MALUKU	4.95	16.01	16.29	13.42	21.09	63.05	10.51	3.25	19.89	16.35	61.71	31.26	7.03
31	MALUKU UTARA	7.80	14.29	18.57	14.64	2.18	65.59	5.29	2.22	10.26	8.57	62.79	30.61	6.59
32	PAPUA BARAT	4.52	13.12	17.05	12.47	27.89	56.73	5.02	1.76	8.05	11.36	64.91	27.78	7.32
33	PAPUA	3.16	15.01	24.60	19.86	25.81	80.73	3.44	1.77	8.44	9.88	88.25	10.41	1.34
INDONESIA		5.83	11.97	13.40	9.66	30.76	53.24	5.94	4.20	10.17	5.78	73.85	20.11	6.05

Lampiran 11 Data Indikator Pasar Tenaga Kerja Sakernas Agustus 2015 (X<sub>1</sub> sampai X<sub>25</sub>)

NO	PROVINSI	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	ACEH	63.44	57.14	35.04	4.45	34.90	9.53	16.08	60.50	44.83	12.73	42.44	26.03
2	SUMATERA UTARA	67.28	62.77	36.81	3.05	34.40	8.47	17.27	60.14	41.30	14.22	44.49	22.98
3	SUMATERA BARAT	64.56	60.11	31.91	3.67	38.77	11.91	13.73	64.41	39.20	13.69	47.11	22.48
4	RIAU	63.22	58.27	46.29	5.07	28.84	9.81	9.99	48.64	42.61	13.41	43.99	25.00
5	JAMBI	66.14	63.27	37.40	3.95	35.36	8.41	14.87	58.65	52.86	10.07	37.07	32.31
6	SUMATERA SELATAN	68.53	64.37	35.41	3.27	35.13	5.43	20.75	61.31	54.74	10.97	34.29	27.69
7	BENGKULU	70.67	67.20	28.06	3.12	39.01	8.82	20.99	68.81	54.21	10.03	35.76	25.27
8	LAMPUNG	65.60	62.23	25.43	3.75	37.66	13.68	19.49	70.83	48.78	16.20	35.02	29.77
9	KEP. BANGKA BELITUNG	66.71	62.51	41.60	4.27	33.34	7.40	13.38	54.13	36.63	22.96	40.40	18.52
10	KEP. RIAU	65.07	61.03	68.49	3.42	21.27	2.71	4.12	28.10	11.03	35.00	53.97	10.74
11	DKI JAKARTA	66.39	61.59	68.16	4.80	19.66	2.65	4.74	27.04	0.42	19.89	79.68	8.18
12	JAWA BARAT	60.34	55.08	46.24	3.37	28.64	14.64	7.11	50.39	16.47	31.09	52.44	14.86
13	JAWA TENGAH	67.86	64.47	34.72	3.56	34.16	14.21	13.35	61.72	28.66	30.15	41.19	20.95
14	DI YOGYAKARTA	68.38	65.60	45.31	3.48	29.58	9.72	11.92	51.21	23.08	23.77	53.15	18.64
15	JAWA TIMUR	67.84	64.81	32.94	3.87	33.78	13.73	15.68	63.19	36.57	22.54	40.89	23.83
16	BANTEN	62.24	56.30	58.58	2.88	22.91	9.46	6.17	38.54	13.02	31.90	55.09	11.88
17	BALI	75.51	74.01	42.95	4.25	29.19	9.85	13.75	52.80	22.40	21.48	56.12	17.97
18	NUSA TENGGARA BARAT	66.54	62.76	24.03	2.77	38.69	19.91	14.60	73.20	39.01	18.52	42.47	20.63
19	NUSA TENGGARA TIMUR	69.25	66.60	19.80	1.61	44.64	3.05	30.90	78.60	61.65	10.52	27.82	38.40
20	KALIMANTAN BARAT	69.68	66.09	31.27	3.17	37.68	5.49	22.40	65.56	57.81	10.42	31.77	26.53
21	KALIMANTAN TENGAH	71.11	67.88	42.26	3.30	33.49	4.82	16.13	54.44	46.32	15.18	38.50	23.31
22	KALIMANTAN SELATAN	69.73	66.30	38.07	3.41	35.86	7.30	15.37	58.52	36.01	16.16	47.83	27.10
23	KALIMANTAN TIMUR	62.39	57.71	56.04	4.30	26.00	6.31	7.35	39.66	22.50	22.86	54.65	14.54
24	SULAWESI UTARA	61.28	55.75	36.41	4.04	34.45	16.36	8.74	59.55	31.93	17.42	50.65	16.88
25	SULAWESI TENGAH	67.51	64.74	28.73	4.64	40.77	7.69	18.16	66.63	50.03	11.59	38.37	25.83
26	SULAWESI SELATAN	60.94	57.31	33.45	3.50	38.59	5.51	18.95	63.05	41.73	13.86	44.41	27.80
27	SULAWESI TENGGARA	68.35	64.56	27.31	3.17	39.30	5.78	24.43	69.52	45.52	14.07	40.41	28.76
28	GORONTALO	63.65	60.69	33.26	3.88	39.50	12.08	11.28	62.86	34.66	16.30	49.04	20.20
29	SULAWESI BARAT	70.27	67.91	23.45	2.87	42.51	6.16	25.00	73.68	58.53	12.20	29.27	39.29
30	MALUKU	64.47	58.07	30.08	1.90	42.59	3.78	21.65	68.02	46.79	9.52	43.69	22.11
31	MALUKU UTARA	66.43	62.41	28.02	3.38	40.39	6.23	21.98	68.60	50.23	11.13	38.64	25.53
32	PAPUA BARAT	68.68	63.13	36.91	2.54	37.64	3.42	19.50	60.55	42.11	10.57	47.32	20.91
33	PAPUA	79.57	76.40	17.56	0.95	42.91	1.44	37.14	81.49	73.93	4.59	21.47	32.38
	INDONESIA	65.76	61.70	38.70	3.54	32.85	10.92	13.99	57.76	32.88	21.84	45.28	21.40

Lampiran 11 (lanjutan)

NO	PROVINSI	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	ACEH	6.65	16.28	17.96	13.60	23.74	56.74	9.93	4.63	17.32	10.99	53.54	34.73	11.73
2	SUMATERA UTARA	5.21	13.09	14.82	11.58	30.82	55.35	6.71	3.51	10.14	9.93	54.15	37.65	8.20
3	SUMATERA BARAT	6.77	12.59	14.53	10.61	26.29	57.72	6.89	3.25	12.63	8.99	64.47	26.95	8.58
4	RIAU	6.73	14.45	13.63	9.86	26.81	43.93	7.83	5.47	11.14	9.51	66.30	27.42	6.28
5	JAMBI	5.32	16.86	20.19	12.72	24.62	54.53	4.34	1.88	9.07	7.20	68.12	25.06	6.81
6	SUMATERA SELATAN	4.42	15.55	18.16	11.56	27.04	57.33	6.07	3.46	12.46	6.35	69.00	24.29	6.71
7	BENGKULU	5.64	15.60	17.05	13.01	24.92	64.54	4.91	2.30	8.28	9.72	61.17	31.97	6.87
8	LAMPUNG	7.37	15.25	15.33	10.76	25.03	64.90	5.14	2.72	11.15	6.90	70.58	24.55	4.87
9	KEP. BANGKA BELITUNG	3.47	10.13	13.47	10.45	36.44	48.73	6.29	3.98	11.10	6.92	70.12	26.37	3.51
10	KEP. RIAU	2.03	5.81	6.43	7.22	43.73	24.23	6.20	4.84	7.82	4.73	57.04	39.72	3.24
11	DKI JAKARTA	2.52	3.51	4.60	5.59	48.96	21.49	7.23	4.80	10.00	5.46	36.87	41.03	22.09
12	JAWA BARAT	4.20	8.86	10.11	8.26	38.90	43.80	8.72	6.70	14.07	5.99	79.71	13.55	6.74
13	JAWA TENGAH	5.56	10.36	11.55	9.22	32.66	54.07	4.99	3.10	10.64	6.06	79.49	15.12	5.38
14	DI YOGYAKARTA	5.99	8.07	8.78	9.05	36.97	44.92	4.07	2.15	6.21	5.26	48.38	32.07	19.55
15	JAWA TIMUR	6.74	12.22	13.29	9.13	30.14	56.41	4.47	2.23	9.94	5.58	76.83	17.12	6.05
16	BANTEN	3.24	6.70	8.43	7.79	47.26	33.46	9.55	8.50	13.00	4.92	77.79	15.71	6.50
17	BALI	3.82	6.98	9.81	8.31	31.64	44.08	1.99	0.82	3.15	3.79	60.51	29.97	9.53
18	NUSA TENGGARA BARAT	8.33	14.82	15.81	10.54	22.15	65.67	5.69	3.93	9.47	7.88	70.66	19.19	10.14
19	NUSA TENGGARA TIMUR	11.05	20.31	19.90	12.87	20.88	74.70	3.83	1.48	9.74	9.64	71.63	22.47	5.90
20	KALIMANTAN BARAT	4.91	14.75	17.26	11.95	3.09	62.18	5.15	3.54	9.58	7.00	78.06	17.10	4.83
21	KALIMANTAN TENGAH	5.93	12.23	14.47	11.79	28.12	49.26	4.54	3.05	8.50	4.53	70.29	24.29	5.43
22	KALIMANTAN SELATAN	7.42	12.74	15.21	11.78	2.46	51.99	4.92	3.28	8.46	6.86	70.55	22.96	6.49
23	KALIMANTAN TIMUR	2.79	6.78	10.87	10.37	33.43	34.77	7.50	5.83	10.59	3.83	61.88	29.15	8.97
24	SULAWESI UTARA	4.05	10.68	11.45	10.61	38.23	54.90	9.03	4.90	15.48	8.73	62.27	32.68	5.04
25	SULAWESI TENGAH	7.79	13.78	15.51	11.20	24.23	60.72	4.10	2.35	8.50	4.09	66.66	24.00	9.34
26	SULAWESI SELATAN	9.28	14.03	14.17	11.39	24.81	57.25	5.95	2.40	12.04	9.19	62.01	24.58	13.40
27	SULAWESI TENGGARA	9.17	13.66	16.36	10.60	21.04	63.38	5.55	2.60	10.45	8.50	54.20	32.45	13.36
28	GORONTALO	5.94	10.91	11.70	11.69	21.24	57.48	4.65	1.81	12.73	5.51	72.73	20.59	6.68
29	SULAWESI BARAT	15.93	19.41	16.77	10.34	17.28	68.78	3.35	1.96	6.30	6.65	60.44	27.54	12.02
30	MALUKU	4.44	13.08	15.63	12.78	26.13	63.55	9.93	2.99	18.11	14.92	56.32	36.98	6.70
31	MALUKU UTARA	7.48	15.41	17.98	14.25	23.41	63.85	6.05	2.64	11.05	9.39	63.13	30.18	6.69
32	PAPUA BARAT	3.39	12.17	17.31	13.00	27.90	55.63	8.08	2.17	16.44	10.95	59.95	26.62	13.43
33	PAPUA	3.37	13.64	27.05	21.67	23.55	79.96	3.99	1.81	11.34	7.85	83.87	14.24	1.88
INDONESIA		5.63	11.37	12.89	9.86	32.22	51.72	6.18	3.77	11.16	6.68	71.08	21.59	7.33

Lampiran 12. Output deteksi *outlier* subset data KILM k<sub>1</sub> 0212 menggunakan *software R*

```
> library(mvoutlier)
> dd.plot(k1_0212, quan=0.9, alpha=0.05)
> mtext(text="Outlier Detection Data k1_0212,quan=0.9 ", side=3, adj=1)
> win.graph()
```

```
$outliers
[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[13] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

$md.cla
[1] 1.7817489 1.7513584 0.7212780 1.9569309 1.7302605 1.2937776 2.6331765
[8] 0.9907171 2.9624539 2.5201704 3.2851783 2.1741070 0.7180783 1.4761366
[15] 0.8460387 2.4344045 3.1787277 1.8270917 2.1130570 1.2595207 2.0835607
[22] 0.8293537 1.9763518 1.4707408 1.0917068 2.0601248 1.0893547 2.2298417
[29] 2.3829008 1.7804579 1.7552338 1.0075040 3.1162775

$md.rob
[1] 1.5636488 1.7407799 0.6241410 1.8056822 1.6974029 1.2146890 2.3185611
[8] 0.8495395 2.5986463 2.5647546 3.7501249 2.1425412 0.6563726 1.3958107
[15] 0.7393047 2.6285937 3.5734894 1.8349607 2.0465598 1.2677939 1.8336598
[22] 0.9442614 2.2401615 1.3734366 1.1033283 1.8445387 0.9696671 2.1055606
[29] 2.4231549 1.6846844 1.7771302 1.0189896 2.7191358
```



Lampiran 13. Provinsi *Outlier* Pada Subset Data  $k_1$  Sakernas Februari 2012 Sampai Sakernas Agustus 2015

PROVINSI	Februari				Agustus			
	2012	2013	2014	2015	2012	2013	2014	2015
1. ACEH								
2. SUMUT								
3. SUMBAR								
4. RIAU								
5. JAMBI								
6. SUMSEL								
7. BENGKULU								
8. LAMPUNG								
9. BABEL								
10. KEP. RIAU					*)			
11. DKI	*)		*)		*)			
12. JABAR								
13. JATENG								
14. DIY								
15. JATIM								
16. BANTEN								
17. BALI	*)			*)				*)
18. NTB							*)	
19. NTT								
20. KALBAR								
21. KALTENG								
22. KALSEL								
23. KALTIM								
24. SULUT								
25. SULTENG								
26. SULSEL								
27. SULTRA								
28. GORONTALO			*)					
29. SULBAR				*)				
30. MALUKU							*)	
31. MALUT								
32. PAPUA BARAT								
33. PAPUA		*)	*)	*)	*)	*)	*)	*)

Keterangan

\*) Provinsi terdeteksi sebagai *outlier* pada *quantile* = 0.9

Lampiran 14. Provinsi *Outlier* Pada Subset Data  $k_2$  Sakernas Februari 2012 Sampai Sakernas Agustus 2015

PROVINSI	Februari				Agustus			
	2012	2013	2014	2015	2012	2013	2014	2015
1. ACEH								
2. SUMUT								
3. SUMBAR								
4. RIAU								
5. JAMBI								
6. SUMSEL								
7. BENGKULU								
8. LAMPUNG								
9. BABEL	*)		*)		*)	*)		
10. KEP. RIAU					*)	*)		
11. DKI	*)	*)	*)	*)	*)	*)	*)	*)
12. JABAR								
13. JATENG			*)					
14. DIY								
15. JATIM								
16. BANTEN								
17. BALI								
18. NTB								
19. NTT								
20. KALBAR								
21. KALTENG								
22. KALSEL								
23. KALTIM								
24. SULUT								
25. SULTENG								
26. SULSEL								
27. SULTRA								
28. GORONTALO								
29. SULBAR								
30. MALUKU								
31. MALUT		*)						
32. PAPUA BARAT								
33. PAPUA								

Keterangan

\*) Provinsi terdeteksi sebagai *outlier* pada *quantile* = 0.9

Lampiran 15. Provinsi *Outlier* Pada Subset Data  $k_3$  Sakernas Februari 2012 Sampai Sakernas Agustus 2015

PROVINSI	Februari				Agustus			
	2012	2013	2014	2015	2012	2013	2014	2015
1. ACEH								
2. SUMUT								
3. SUMBAR								
4. RIAU								
5. JAMBI								
6. SUMSEL								
7. BENGKULU								
8. LAMPUNG								
9. BABEL	*)	*)						
10. KEP. RIAU	*)				*)	*)		
11. DKI			*)				*)	
12. JABAR								
13. JATENG								
14. DIY		*)						
15. JATIM								
16. BANTEN			*)					*)
17. BALI								*)
18. NTB		*)					*)	
19. NTT				*)				
20. KALBAR								
21. KALTENG								
22. KALSEL						*)		
23. KALTIM			*)					
24. SULUT								
25. SULTENG								
26. SULSEL								
27. SULTRA								
28. GORONTALO						*)		
29. SULBAR								
30. MALUKU	*)			*)	*)		*)	
31. MALUT								
32. PAPUA BARAT								
33. PAPUA				*)	*)			*)

Keterangan

\*) Provinsi terdeteksi sebagai *outlier* pada *quantile* = 0.9

Lampiran 16. Provinsi *Outlier* Pada Subset Data  $k_4$  Sakernas Februari 2012 Sampai Sakernas Agustus 2015

PROVINSI	Februari				Agustus			
	2012	2013	2014	2015	2012	2013	2014	2015
1. ACEH								
2. SUMUT								
3. SUMBAR								
4. RIAU								
5. JAMBI								
6. SUMSEL								
7. BENGKULU								
8. LAMPUNG								
9. BABEL								
10. KEP. RIAU								
11. DKI	*)	*)			*)	*)	*)	
12. JABAR								
13. JATENG			*)					
14. DIY	*)							
15. JATIM	*)							
16. BANTEN		*)						
17. BALI								
18. NTB								
19. NTT								
20. KALBAR			*)					*)
21. KALTENG								
22. KALSEL								*)
23. KALTIM								
24. SULUT								
25. SULTENG								
26. SULSEL								
27. SULTRA								
28. GORONTALO								
29. SULBAR		*)	*)	*)	*)	*)		
30. MALUKU				*)				
31. MALUT							*)	
32. PAPUA BARAT								
33. PAPUA				*)	*)	*)	*)	*)

Keterangan

\*) Provinsi terdeteksi sebagai *outlier* pada *quantile* = 0.9

Lampiran 17. Provinsi *Outlier* Pada Subset Data  $k_5$  Sakernas Februari 2012 Sampai Sakernas Agustus 2015

PROVINSI	Februari				Agustus			
	2012	2013	2014	2015	2012	2013	2014	2015
1. ACEH								
2. SUMUT								
3. SUMBAR								
4. RIAU								
5. JAMBI								
6. SUMSEL								
7. BENGKULU								
8. LAMPUNG								
9. BABEL								
10. KEP. RIAU	*)	*)				*)	*)	
11. DKI	*)	*)		*)	*)		*)	
12. JABAR								
13. JATENG								
14. DIY								
15. JATIM								
16. BANTEN								
17. BALI	*)	*)	*)	*)	*)	*)		*)
18. NTB								
19. NTT								
20. KALBAR			*)					
21. KALTENG								
22. KALSEL								
23. KALTIM								
24. SULUT								
25. SULTENG								
26. SULSEL								
27. SULTRA								
28. GORONTALO								
29. SULBAR								
30. MALUKU								
31. MALUT								
32. PAPUA BARAT								
33. PAPUA			*)	*)	*)	*)	*)	*)

Keterangan

\*) Provinsi terdeteksi sebagai *outlier* pada *quantile* = 0.9

Lampiran 18. *Output Mardia test* Untuk Uji Normal Multivariat Subset Data  
KILM k<sub>1</sub> 0212 – k<sub>1</sub> 0815

```
> library(psych)
> mardia(k1_0212)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 5.28  skew = 29.05  with probability = 0.087
small sample skew = 32.85  with probability = 0.035
b2p = 23.02  kurtosis = -0.41  with probability = 0.68

> mardia(k1_0213)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 4.98  skew = 27.39  with probability = 0.12
small sample skew = 30.98  with probability = 0.055
b2p = 22.33  kurtosis = -0.69  with probability = 0.49

> mardia(k1_0214)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 4.81  skew = 26.44  with probability = 0.15
small sample skew = 29.9  with probability = 0.071
b2p = 23.89  kurtosis = -0.05  with probability = 0.96

> mardia(k1_0215)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 4.58  skew = 25.2  with probability = 0.19
small sample skew = 28.49  with probability = 0.098
b2p = 25.58  kurtosis = 0.65  with probability = 0.51

> mardia(k1_0812)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 5.4  skew = 29.72  with probability = 0.074
small sample skew = 33.61  with probability = 0.029
b2p = 25.91  kurtosis = 0.79  with probability = 0.43

> mardia(k1_0813)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 5.39  skew = 29.64  with probability = 0.076
small sample skew = 33.52  with probability = 0.03
b2p = 23.79  kurtosis = -0.09  with probability = 0.93

> mardia(k1_0814)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 8.26  skew = 45.44  with probability = 0.00096
small sample skew = 51.39  with probability = 0.00014
b2p = 28.97  kurtosis = 2.06  with probability = 0.039

> mardia(k1_0815)
Mardia tests of multivariate skew and kurtosis
Use describe(x) the to get univariate tests
n.obs = 33  num.vars = 4
b1p = 4.73  skew = 26.03  with probability = 0.16
small sample skew = 29.43  with probability = 0.08
b2p = 23.94  kurtosis = -0.03  with probability = 0.98
```

Lampiran 19. Hasil Uji Normal Multivariat Subset Data KILM k<sub>1</sub> 0215 – k<sub>5</sub> 0815

Kombinasi	<u>Sakernas Februari</u>				<u>Sakernas Agustus</u>			
	2012	2013	2014	2015	2012	2013	2014	2015
k <sub>1</sub>	*	**) )	**) )	**) )	*	*	*	**) )
k <sub>2</sub>	*	*	*	*	*	*	*	*
k <sub>3</sub>	*	*	**) )	*	*	*	*	*
k <sub>4</sub>	*	*	*	*	*	*	*	*
k <sub>5</sub>	*	**) )	*	*	*	*	*	*

Keterangan :

- \* : data tidak berdistribusi normal multivariat, maka asumsi data berdistribusi  $t$  multivariat terpenuhi
- \*\*) : data berdistribusi normal multivariat, maka asumsi data berdistribusi  $t$  multivariat tidak terpenuhi

Lampiran 20. *Script dan Output Kelompok Optimal Subset Data KILM k<sub>2</sub> 0212 – k<sub>2</sub> 0815 dengan Kriteria ICL Menggunakan Software R*

```
> teigen(k2_0212)$iclresult$classification
[1] 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2

> teigen(k2_0213)$iclresult$classification
[1] 2 2 2 2 2 2 2 2 1 1 2 1 1 2 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2

> teigen(k2_0214)$iclresult$classification
[1] 1 1 1 1 1 1 1 1 2 2 1 2 2 1 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1

> teigen(k2_0215)$iclresult$classification
[1] 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2

> teigen(k2_0812)$iclresult$classification
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

> teigen(k2_0813)$iclresult$classification
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

> teigen(k2_0814)$iclresult$classification
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

> teigen(k2_0815)$iclresult$classification
[1] 2 2 2 2 2 2 2 2 1 1 2 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```



Lampiran 21. Output Nilai ICL Subset Data KILM k<sub>2</sub> 0212

	G=1	G=2	G=3	G=4	G=5	G=6	G=7	G=8	G=9
UUUU	210.8397	207.3997	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUUC	-Inf	210.9106	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCU	-Inf	204.3836	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCC	-Inf	208.0415	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUU	-Inf	204.4373	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUC	-Inf	207.7625	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCU	-Inf	208.3708	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCC	-Inf	205.4104	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUU	-280.2091	-246.8512	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUC	-Inf	-244.7098	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CICU	-Inf	-240.5697	-241.5384	-238.3051	-Inf	-Inf	-Inf	-Inf	-Inf
CICC	-Inf	-238.3925	-238.4017	-234.7289	-Inf	-Inf	-Inf	-Inf	-Inf
UIIU	-274.2884	-235.7203	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIIC	-Inf	-232.5057	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIU	-Inf	-236.8258	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIC	-Inf	-234.4657	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUU	-Inf	-246.2012	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUC	-Inf	-242.8813	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCU	-Inf	207.0865	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCC	-Inf	208.0909	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCU	-Inf	201.0824	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCC	-Inf	204.5773	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICU	-Inf	-239.9150	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICC	-Inf	-236.7318	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUU	-Inf	-256.6955	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUC	-Inf	-253.3724	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUU	-Inf	-257.3456	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUC	-Inf	-255.2042	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf

\$bestmodel  
[1] "The best model (ICL of 210.91) is UUUC with G=2"

\$classification  
[1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1

Lampiran 22. Estimasi Parameter Subset Data KILM k<sub>2</sub> 0212 Pada Jumlah Kelompok (**G**) = 2 , Model UUUC Menggunakan MBC- ICL

```

$parameters$df
[1] 52.0468 52.0468

$parameters$mean
      [,1]      [,2]      [,3]
[1,] 0.6274744 -0.679212 -0.4845030
[2,] -0.9976486 1.083394 0.7681694

$parameters$weights
      [,1]      [,2]
[1,] 1.0472011 0.8530383
[2,] 0.9277504 0.7069291
[3,] 0.9691905 0.8893973
[4,] 0.9713066 0.8453603
[5,] 0.9852402 0.7479034
[6,] 1.0251542 0.7759443
[7,] 1.0349316 0.7689999
[8,] 1.0463489 0.8653294
[9,] 0.2882972 0.9562541
[10,] 0.7010253 1.0016068
[11,] 0.4696034 0.9157909
[12,] 0.4190115 1.0044022
[13,] 0.4631573 1.0364873
[14,] 0.7626971 1.0507933
[15,] 0.7757501 0.9734907
[16,] 0.5240723 1.0110464
[17,] 0.7720868 1.0151602
[18,] 1.0372142 0.9161978
[19,] 0.9848767 0.6995370
[20,] 0.9477970 0.7730767
[21,] 0.9674087 0.8381739
[22,] 0.6424582 1.0302466
[23,] 0.8009369 1.0471400
[24,] 0.9316752 1.0119393
[25,] 0.9692115 0.9187202
[26,] 1.0125776 0.8916007
[27,] 1.0058795 0.8583729
[28,] 0.5718538 0.9591844
[29,] 1.0197070 0.7559765
[30,] 1.0529467 0.8377534
[31,] 1.0545560 0.8152212
[32,] 1.0475452 0.8871134
[33,] 0.9171549 0.6147770

$parameters$sigma
, , 1
      [,1]      [,2]      [,3]
[1,] 0.2548055 -0.1549152 -0.2716632
[2,] -0.1549152 0.1326448 0.1412740
[3,] -0.2716632 0.1412740 0.3044764
, , 2
      [,1]      [,2]      [,3]
[1,] 0.42742722 -0.03853226 -0.5929850
[2,] -0.03853226 0.37517740 -0.1770302
[3,] -0.59298503 -0.17703019 0.9655910

$parameters$piG
[1] 0.6136783 0.3863217

```

Lampiran 23. Pengelompokan Provinsi di Indonesia Menurut Kombinasi Variabel  $k_2$  (Persentase Bekerja Menurut Lapangan Usaha) Pada Jumlah Kelompok ( $G$ ) = 2 Menggunakan MBC- ICL

PROVINSI	0212	0213	0214	0215	0812	0813	0814	0815
1. ACEH	1	1	1	1	1	1	1	1
2. SUMUT	1	1	1	1	1	1	1	1
3. SUMBAR	1	1	1	1	1	1	1	1
4. RIAU	1	1	1	1	1	1	1	1
5. JAMBI	1	1	1	1	1	1	1	1
6. SUMSEL	1	1	1	1	1	1	1	1
7. BENGKULU	1	1	1	1	1	1	1	1
8. LAMPUNG	1	1	1	1	1	1	1	1
9. BABEL	2	2	2	2	2	2	2	2
10. KEP. RIAU	2	2	2	2	2	2	2	2
11. DKI	2	1	1	2	2	1	2	1
12. JABAR	2	2	2	2	2	2	2	2
13. JATENG	2	2	2	2	2	2	2	2
14. DIY	2	1	1	2	1	1	2	1
15. JATIM	2	2	2	2	1	1	2	2
16. BANTEN	2	2	2	2	2	2	2	2
17. BALI	2	1	2	2	1	1	2	1
18. NTB	1	1	1	1	1	1	1	1
19. NTT	1	1	1	1	1	1	1	1
20. KALBAR	1	1	1	1	1	1	1	1
21. KALTENG	1	1	1	1	1	1	1	1
22. KALSEL	2	2	2	1	1	1	1	1
23. KALTIM	2	1	1	2	1	1	2	1
24. SULUT	2	1	1	1	1	1	1	1
25. SULTENG	1	1	1	1	1	1	1	1
26. SULSEL	1	1	1	1	1	1	1	1
27. SULTRA	1	1	1	1	1	1	1	1
28. GORONTALO	2	1	1	1	1	1	1	1
29. SULBAR	1	1	1	1	1	1	1	1
30. MALUKU	1	1	1	1	1	1	1	1
31. MALUT	1	1	1	1	1	1	1	1
32. PAPUA BARAT	1	1	1	1	1	1	1	1
33. PAPUA	1	1	1	1	1	1	1	1

Lampiran 24. Output Nilai ICL Subset Data KILM k<sub>5</sub> 0815

\$allicl	G=1	G=2	G=3	G=4	G=5	G=6	G=7	G=8	G=9
UUUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCU	-Inf	233.6887	215.9113	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCC	-Inf	237.1732	222.8063	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCU	-Inf	239.0902	219.9844	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCC	-Inf	243.6913	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUU	-278.0484	-260.8217	-251.4207	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUC	-Inf	-255.8055	-239.3129	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CICU	-Inf	-257.1459	-242.2674	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CICC	-Inf	-253.6240	-237.6734	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIIU	-271.2482	-254.7225	-244.7425	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIIC	-Inf	-251.2508	-239.5599	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIU	-Inf	-251.2438	-242.7314	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIC	-Inf	-247.7502	-235.4470	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUU	-Inf	-267.3233	-250.1568	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUC	-Inf	-263.4991	-243.1278	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCU	-Inf	231.7247	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCC	-Inf	235.1333	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCU	-Inf	222.6445	190.3437	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCC	-Inf	226.1195	197.2925	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICU	-Inf	-261.5888	-250.1475	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICC	-Inf	-257.3668	-243.2062	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUU	-Inf	-277.8977	-260.6283	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUC	-Inf	-274.1090	-253.6422	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUU	-Inf	-271.3056	-261.9102	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUC	-Inf	-266.3331	-249.7800	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf

Lampiran 25. Pengelompokan Provinsi di Indonesia Menurut Kombinasi Variabel  $k_3$  (Bekerja Menurut Status, dan TPT Menurut Pendidikan) Pada Jumlah Kelompok Optimal Menggunakan MBC-ICL

PROVINSI	0212	0213	0214	0215	0812	0813	0814	0815
1. ACEH	1	2	2	2	2	2	2	2
2. SUMUT	2	2	2	2	2	2	2	2
3. SUMBAR	2	2	2	2	2	2	2	2
4. RIAU	1	2	2	2	2	2	2	1
5. JAMBI	2	2	2	2	2	2	2	2
6. SUMSEL	2	2	2	2	2	2	2	2
7. BENGKULU	2	2	2	2	2	2	2	2
8. LAMPUNG	2	2	2	2	2	2	2	2
9. BABEL	1	2	2	2	2	2	2	2
10. KEP. RIAU	1	1	1	1	2	2	1	1
11. DKI	1	1	1	1	1	1	1	1
12. JABAR	2	1	1	1	1	1	1	1
13. JATENG	2	2	2	2	2	2	2	2
14. DIY	2	2	2	2	2	2	2	2
15. JATIM	2	2	2	2	2	2	2	2
16. BANTEN	1	1	1	1	1	1	1	1
17. BALI	2	2	2	2	2	2	2	2
18. NTB	2	2	2	2	2	2	2	2
19. NTT	2	2	2	2	2	2	2	2
20. KALBAR	2	2	2	2	2	2	2	2
21. KALTENG	2	2	2	2	2	2	2	2
22. KALSEL	2	2	2	2	2	2	2	2
23. KALTIM	1	1	1	1	1	1	1	1
24. SULUT	2	2	2	2	2	2	2	2
25. SULTENG	2	2	2	2	2	2	2	2
26. SULSEL	2	2	2	2	2	2	2	2
27. SULTRA	2	2	2	2	2	2	2	2
28. GORONTALO	2	2	2	2	2	2	2	2
29. SULBAR	2	2	2	2	2	2	2	2
30. MALUKU	2	2	2	2	2	2	1	2
31. MALUT	2	2	2	2	2	2	2	2
32. PAPUA BARAT	2	2	2	2	2	2	2	2
33. PAPUA	2	2	2	2	2	2	2	2

Lampiran 26. Output Nilai ICL Subset Data KILM k<sub>3</sub> 0815

	G=1	G=2	G=3	G=4	G=5	G=6	G=7
UUUU	-245.4874	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCU	-Inf	-260.5112	-264.5170	-285.7740	-290.9559	-301.7623	-331.8589
CCCC	-Inf	-257.0477	-257.6413	-275.4215	-277.4015	-284.7699	-317.4791
CIUU	-753.8229	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CICU	-Inf	-717.2611	-709.0588	-715.3982	-704.3200	-711.5942	-713.9137
CICC	-Inf	-714.2741	-702.8846	-707.0407	-693.6756	-699.1277	-Inf
UIIU	-735.2895	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIIC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
	G=8	G=9					
UUUU	-Inf	-Inf					
UUUC	-Inf	-Inf					
CUCU	-Inf	-Inf					
CUCC	-Inf	-Inf					
CUUU	-Inf	-Inf					
CUUC	-Inf	-Inf					
CCCU	-334.9829	-341.9461					
CCCC	-327.3363	-313.6741					
CIUU	-Inf	-Inf					
CIUC	-Inf	-Inf					
CICU	-727.5816	-730.0415					
CICC	-707.5450	-707.0495					
UIIU	-Inf	-Inf					
UIIC	-Inf	-Inf					
CIIU	-Inf	-Inf					
CIIC	-Inf	-Inf					
UIUU	-Inf	-Inf					
UIUC	-Inf	-Inf					
UCCU	-Inf	-Inf					
UCCC	-Inf	-Inf					
UUCU	-Inf	-Inf					
UUCC	-Inf	-Inf					
UICU	-Inf	-Inf					
UICC	-Inf	-Inf					
UCUU	-Inf	-Inf					
UCUC	-Inf	-Inf					
CCUU	-Inf	-Inf					
CCUC	-Inf	-Inf					

Lampiran 27. Pengelompokan Provinsi di Indonesia Menurut Kombinasi Variabel  $k_4$  (Persentase Bekerja Menurut Jam Kerja, dan Persentase Pekerja Setengah Penganggur) Pada Jumlah Kelompok ( $G$ ) = 2 Menggunakan MBC- ICL

PROVINSI	0212	0213	0214	0215	0812	0813	0814	0815
1. ACEH	1	1	1	1	1	1	1	1
2. SUMUT	1	1	1	1	2	2	2	1
3. SUMBAR	1	1	1	1	1	1	1	1
4. RIAU	1	1	1	1	1	1	1	1
5. JAMBI	1	1	1	1	1	1	1	1
6. SUMSEL	1	1	1	1	1	1	1	1
7. BENGKULU	1	2	2	1	1	1	1	1
8. LAMPUNG	1	1	1	1	1	1	1	1
9. BABEL	2	2	2	1	2	1	2	2
10. KEP. RIAU	2	2	2	2	2	2	2	2
11. DKI	2	2	2	2	2	2	2	1
12. JABAR	2	2	2	1	1	1	1	2
13. JATENG	2	2	2	1	1	1	1	2
14. DIY	2	2	2	1	2	2	2	1
15. JATIM	2	1	1	1	1	1	1	2
16. BANTEN	2	2	2	2	1	2	2	2
17. BALI	1	1	1	1	2	1	2	2
18. NTB	1	1	1	1	1	1	1	1
19. NTT	1	1	1	1	1	1	1	1
20. KALBAR	1	1	1	1	1	1	1	1
21. KALTENG	1	2	2	1	1	1	1	1
22. KALSEL	1	1	1	1	1	1	1	1
23. KALTIM	1	2	2	1	1	1	2	2
24. SULUT	1	2	2	1	2	2	2	2
25. SULTENG	1	1	1	1	1	1	1	1
26. SULSEL	1	1	1	1	1	1	1	1
27. SULTRA	1	1	1	1	1	1	1	1
28. GORONTALO	1	1	1	1	1	1	1	1
29. SULBAR	1	1	1	1	1	1	1	1
30. MALUKU	1	1	1	1	2	1	1	1
31. MALUT	1	1	1	1	1	1	1	1
32. PAPUA BARAT	1	2	2	1	1	1	1	1
33. PAPUA	1	1	1	1	1	1	1	1

Lampiran 28. Output Nilai ICL Subset Data KILM k<sub>4</sub> 0815

\$allicl	G=1	G=2	G=3	G=4	G=5	G=6	G=7	G=8	G=9
UUUU	-184.3677	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUCC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CUUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCU	-Inf	-205.3089	-208.9090	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCCC	-Inf	-207.3735	-209.4339	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUU	-734.1401	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CICU	-Inf	-704.1209	-706.8202	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CICC	-Inf	-700.6200	-705.4683	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIIU	-712.2395	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIIC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CIIC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UIUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCCC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UUCC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UICC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
UCUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUU	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
CCUC	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf



Lampiran 29. *Script Model-Based Clustering Dengan Kriteria ICL Pada Software R*

```
# Memanggil data
>prov0212<-read.table (file.choose(), header=T, sep="\t")
>data0212= data.frame(prov0212[ ,2:26])

# Membentuk kombinasi variabel k2 yang terdiri dari data pada kolom 9,
10 dan 11.
>k2_0212= as.matrix(data0212[ ,c(9,10,11)])

# Deteksi outlier subset data k2_0212
>library(mvoutlier)
>dd.plot(k2_0212, quan=0.9, alpha=0.05)
>mtext(text="Outlier Detection Data k2_0212,quan=0.9 ", side=3, adj=1)

# Uji Asumsi Normal Multivariat subset data k2_0212
>library(psych)
>mardia(k2_0212)

# Klasifikasi dan estimasi parameter kelompok optimal dengan kriteria ICL
>library(teigen)
>teigen(k2_0212)$iclresult

# Menampilkan Anggota kelompok Model UUUC dengan jumlah kelompok G=2
>library(teigen)
>teigen(k2_0212)
>tk2_0212_UUUC <-teigen(k2_0212, models="UUUC",parallel.cores=2, Gs=2)
>tk2_0212_UUUC$iclresults$classification

# Marginal Contour Plot subset data k2_0212
>library(teigen)
>tk20212 <- teigen(k2_0212, models="UUUC", parallel.cores=2, Gs=2,
init="hard")
>plot(tk20212, what = "contour")
>mtext(text="Subset Data k2_0212", side=3, adj=1)
```

## Lampiran 30 Program RMBC-MML

```
% -----
%
% ROBUST MODEL-BASED CLUSTERING WITH MULTIVARIATE t
% DISTRIBUTIONS AND MINIMUM MESSAGE LENGTH (RMBC-MML)
%
% Source      : private communication
% Author      : Associate Professor Shy Shoham
% Department of Biomedical Engineering,
% Technion, Israel
% Modification : Tiodora H.Siagian, M.Pop.Hum.Res
% PhD Student, Statistics Department, ITS,
% Surabaya, Pebruari 2014
%
% -----

function
[bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=RMBCmml(data,Pj,mu,Sigma,n
u,N,max_iteration,method)
% This function iteratively calculate the estimation of a
% multivariate t
% mixture model using EM algorithm and then cluster the data
%
% Usage syntax:
% [bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=
RMBCmml(data,Pj,mu,Sigma,nu,N,max_iteration,method);
%
% Inputs:
% data      : the data; for n observations in p dimensions
% Pj        : the initial mixing proportion
% mu        : the initial mean vector from kmeans result
% Sigma     : the matrix identity
% nu        : a scalar specifies the degree of freedom
% N         : number of parameter in a cluster
% method    : has 2 options: 'agglomerate' and 'regular'
%
% Outputs:
% bestPj     : estimated mixing proportion
% bestmu     : estimated mean vectors
% bestSigma  : estimated covariance matrix
% nu         : degree of freedom
% indZ       : cluster which has max Z
% Z          : estimated posterior probabilities
% U          : weight
% M          : Mahalanobis distance

%Initialization

format compact;
[n,d]=size(data);% find the size of data

%plot stuff
```

```

set(gcf,'CurrentAxes','PCA');
cla
result = robpca(data);
hold on

plot(result.T(:,1),result.T(:,2),'k.','MarkerSize',12);
xlabel('PC 1'),ylabel('PC 2');
ph1=plot(0,0);
colors= repmat(['b'],20,1);

g=length(Pj);
if g>1
    Pj=sparse(Pj);
end
rep=reshape(repmat(1:g,n,1),g*n,1);
rep_data=repmat(data,g,1);
diffs=rep_data-mu(rep,:); %Distances to cluster center
delta=1;
Lmax=-inf; %Initial likelihood
L=Lmax;
nu_old=nu;
Lhist=[]
ghist=[]
SigmaNo=length(Sigma);
detSigma=sparse(zeros(1,SigmaNo));
for i=1:SigmaNo
    detSigma(i)=1/sqrt(det(Sigma{i}));
end
M=zeros(n,g);
for i=1:SigmaNo
    i
    M(:,i)=sum(((sqrtm(pinv(Sigma{i}))*diffs(rep==i,:)).^2));
%Mahalanobis distances
end
c=gamma((nu+d)/2)/(gamma(nu/2)*(pi*nu)^(d/2))
Prob=[c*exp(-(nu+d)*log(1+M/nu)/2)*diag(detSigma)] %Pdf
multivariate t
k=0 %iteration counter

% Start Algorithm

while (g>=1)&&(k<max_iteration)
    while ((delta>0.1)||abs(nu-nu_old)>1e-1)&&(k<max_iteration)
        k=k+1;
        %plot ellipses
        delete(ph1)
        ph1=[];

        for i=1:g
            sig1=result.P(:,1:2)'+(Sigma{i}*result.P(:,1:2));%rotate
covariance to pc dimensions
            [V,D]=eig(sig1);
            center=(mu(i,:)-mean(data))*result.P(:,1:2);

```

```

        %angle -angle(Z) returns the phase angles, in radians,
for each
        %element of complex array Z. The angles lie between +/-
phi
        if Pj(i)>0

ph1(i) =
ellipse(2*sqrt(D(1,1)),2*sqrt(D(2,2)),angle(V(1,1)+1i*V(2,1)),cent
er(1),center(2),colors(i+2));
        set(ph1(i),'LineWidth',Pj(i)*20);
        else

        ph1(i)=plot(center(1),center(2),'EraseMode','xor');
        end
    end
drawnow

%------%
%           E step           %
%------%

U=(nu+d)./(nu+M);
%temp=Prob*diag(Pj);
%Z=diag(sparse(1./sum(temp,2)))*temp;
Z=Prob*diag(Pj)./(sum(Prob*diag(Pj),2)*ones(1,g));
ZU=Z.*U;

%------%
%           M step           %
%------%

switch method
case 'agglomerate'
    deltaP=1;
    gtemp=g;
    while deltaP>10^-4
        Pjold=Pj;
        temp1=Prob*diag(Pj);
        temp2=sum(temp1,2);
        for j=1:g
            if Pj(j)>0
                Pj(j)=max((sum(temp1(:,j))./temp2)-N/2)/(n-
N/2*gtemp),0);
            if Pj(j)==0
                temp2=sum(Prob*diag(Pj),2);
                gtemp=gtemp-1;%number of clusters
            end
        end
        end
        deltaP=norm(Pj-Pjold,1);
    end
case 'regular'

Pj=sparse(sum(Prob*diag(Pj)./(sum(Prob*diag(Pj),2)*ones(1,g)))/n);
end

```

```

mu=(ZU'*data)./(sum(ZU)'*ones(1,d))%

%Update DOF parameter
y=-sum(sum(Z.*(psi(0,(nu+d)/2)+log(2./(nu+M))-
U)))/sum(sum(Z));
temp=1/(y+log(y)-1);
nu_old=nu;
nu=min(2*temp+0.0416*(1+erf(0.6594*log(2.1971*temp))),100);

%Calculate Covariance
detSigma=[];
for i=1:g
    diffs=data-ones(n,1)*mu(i,:);

Sigma{i}=(((ZU(:,i)*ones(1,d)).*diffs)'*diffs)/sum(ZU(:,i));
    detSigma(i)=sparse(1/sqrt(det(Sigma{i})));
    M(:,i)=sum(((sqrtm(pinv(Sigma{i}))*diffs').^2))';
%Mahalanobis distances
end
% update Pdf multivariate t
c=gamma((nu+d)/2)/(gamma(nu/2)*(pi*nu)^(d/2));
Prob=[c*exp(-(nu+d)*log(1+M/nu)/2)*diag(detSigma)];
switch method
case 'agglomerate'
    % cycle is done, purge empty clusters
    tokeep=find(Pj);
    if length(tokeep)<g
        g=length(tokeep);
        rep=reshape(repmat(1:g,n,1),g*n,1);
        %rep_data=repmat(data,g,1);
        Pj=Pj(tokeep);
        mu=mu(tokeep,:);
        Prob=Prob(:,tokeep);
        Z=Z(:,tokeep);
        U=U(:,tokeep);
        ZU=ZU(:,tokeep);
        M=M(:,tokeep);
        Sigma=Sigma([tokeep]);
    end
    % Update and compare likelihood
    oldL=L;
    L=sum(log(sum(Prob*diag(sparse(Pj)),2)))-
[N/2*sum(log(n*Pj/12))+g/2*log(n/12)+g*(N+1)/2];%MML penalty

    Lhist=[Lhist L];
    ghist=[ghist g];
    delta=abs(L-oldL);
end
end
switch method
case 'agglomerate'

    if L>Lmax
        Lmax=L;%current likelihood is optimal
        Pjopt=Pj;%store optimal parameters

```

```

muopt=mu;
Sigma_optimal=Sigma;
nuopt=nu;
Zopt=Z;
Uopt=U;
Mopt=M;
else
break
end

%purge smallest cluster
[Y,ind]=min(Pj); %find minimal mixing proportion
temp_I=1:g;
disp(['eliminating'])
I=setdiff(temp_I,ind);%find set difference of two vectors

Pj=Pj(I);
mu=mu(I,:);
Prob=Prob(:,I);
Z=Z(:,I);
U=U(:,I);
ZU=ZU(:,I);
M=M(:,I);
Sigma=Sigma([I]);
g=g-1;
rep=reshape(repmat(1:g,n,1),g*n,1);
rep_data=repmat(data,g,1);
delta=1;
end
end

switch method
case 'agglomerate'
bestmu=muopt;
bestSigma=Sigma_optimal;
bestPj=Pjopt;
nu=nuopt;
Z=Zopt;
U=Uopt;
M=Mopt;
[MaxZ indZ]=max(Z');
end

%-----%
%          ellipse          %
%-----%

function h=ellipse(ra,rb,ang,x0,y0,C,Nb)
% Ellipse adds ellipses to the current plot
%
% ELLIPSE(ra,rb,ang,x0,y0) adds an ellipse with semimajor axis of
ra,
% a semimajor axis of radius rb, a semimajor axis of ang, centered
at
% the point x0,y0.

```

```

%
% The length of ra, rb, and ang should be the same.
% If ra is a vector of length L and x0,y0 scalars, L ellipses
% are added at point x0,y0.
% If ra is a scalar and x0,y0 vectors of length M, M ellipse are
% with the same
% radii are added at the points x0,y0.
% If ra, x0, y0 are vectors of the same length L=M, M ellipses are
% added.
% If ra is a vector of length L and x0, y0 are vectors of length
% M~L, L*M ellipses are added, at each point x0,y0, L ellipses of
% radius ra.
%
% ELLIPSE(ra,rb,ang,x0,y0,C)
% adds ellipses of color C. C may be a string ('r','b',...) or the
% RGB value.
% If no color is specified, it makes automatic use of the colors
% specified by
% the axes ColorOrder property. For several circles C may be a
% vector.
%
% ELLIPSE(ra,rb,ang,x0,y0,C,Nb), Nb specifies the number of points
% used to draw the ellipse. The default value is 300. Nb may be
% used
%
% for each ellipse individually.
%
% h=ELLIPSE(...) returns the handles to the ellipses.
%
% as a sample of how ellipse works, the following produces a red
% ellipse
% tipped up at a 45 deg axis from the x axis
% ellipse(1,2,pi/8,1,1,'r')
%
% note that if ra=rb, ELLIPSE plots a circle
%
%
% written by D.G. Long, Brigham Young University, based on the
% CIRCLES.m original
% written by Peter Blattner, Institute of Microtechnology,
% University of
% Neuchatel, Switzerland, blattner@imt.unine.ch

if nargin<1,
    ra=[];
end;
if nargin<2,
    rb=[];
end;
if nargin<3,
    ang=[];
end;

%if nargin==1,
%    error('Not enough arguments');
%end;

```

```

if nargin<5,
    x0=[];
    y0=[];
end;

if nargin<6,
    C=[];
end

if nargin<7,
    Nb=[];
end

% set up the default values

if isempty(ra),ra=1;end;
if isempty(rb),rb=1;end;
if isempty(ang),ang=0;end;
if isempty(x0),x0=0;end;
if isempty(y0),y0=0;end;
if isempty(Nb),Nb=300;end;
if isempty(C),C=get(gca,'colororder');end;

% work on the variable sizes

x0=x0(:);
y0=y0(:);
ra=ra(:);
rb=rb(:);
ang=ang(:);
Nb=Nb(:);

if ischar(C),C=C(:);end;

if length(ra)~=length(rb),
    error('length(ra)~=length(rb)');
end;
if length(x0)~=length(y0),
    error('length(x0)~=length(y0)');
end;

% how many inscribed ellipses are plotted

if length(ra)~=length(x0)
    maxk=length(ra)*length(x0);
else
    maxk=length(ra);
end;

% drawing loop

for k=1:maxk

```



```

if length(x0)==1
    xpos=x0;
    ypos=y0;
    radm=ra(k);
    radn=rb(k);
    if length(ang)==1
        an=ang;
    else
        an=ang(k);
    end;
elseif length(ra)==1
    xpos=x0(k);
    ypos=y0(k);
    radm=ra;
    radn=rb;
    an=ang;
elseif length(x0)==length(ra)
    xpos=x0(k);
    ypos=y0(k);
    radm=ra(k);
    radn=rb(k);
    an=ang(k);
else

    rada=ra(fix((k-1)/size(x0,1))+1);
    radb=rb(fix((k-1)/size(x0,1))+1);
    an=ang(fix((k-1)/size(x0,1))+1);
    xpos=x0(rem(k-1,size(x0,1))+1);
    ypos=y0(rem(k-1,size(y0,1))+1);
end;

co=cos(an);
si=sin(an);
the=linspace(0,2*pi,Nb(rem(k-1,size(Nb,1))+1,:)+1);
% x=radm*cos(the)*co-si*radn*sin(the)+xpos;
% y=radm*cos(the)*si+co*radn*sin(the)+ypos;
h(k)=line(radm*cos(the)*co-
si*radn*sin(the)+xpos,radm*cos(the)*si+co*radn*sin(the)+ypos);
set(h(k),'color',C(rem(k-1,size(C,1))+1,:));
end;

```

Lampiran 31. *Main Script* RMBC-MML Kombinasi Variabel  $k_1$ ,  $k_2$ ,  $k_3$ ,  $k_4$  dan  $k_5$   
 Pada *Software* MATLAB, Sakernas Februari 2012

```
Pj=[0.25,0.25,0.25,0.25];%inisial mixing proportion
p=4;%jumlah variabel
N=9;%jumlah parameter per cluster
nu=9;% inisial derajat bebas
 [~,mu,~]=kmeans(k1_0212,p); %inisialisasi rata-rata menggunakan k-
meansc
for j=1:p
    Sigma{j}=eye(p); %inisialisasi matriks Sigma= matiriks identitas
    ukuran pxp sebanyak p
end

[bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=RMBCmml(k1_0212,Pj,mu,Sigm
a,nu,N,5000,'agglomerate')
```

```
Pj=[0.3,0.3,0.4];%inisial mixing proportion
p=3;%jumlah variabel
N=7;%jumlah parameter per cluster
nu=7;% inisial derajat bebas
 [~,mu,~]=kmeans(k2_0212,p); %inisialisasi rata-rata menggunakan k-
meansc
for j=1:p
    Sigma{j}=eye(p); %inisialisasi matriks Sigma= matiriks identitas
    ukuran pxp sebanyak p
end

[bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=RMBCmml(k2_0212,Pj,mu,Sigm
a,nu,N,5000,'agglomerate')
```

```
Pj=[0.125,0.125,0.125,0.125,0.125,0.125,0.125,0.125];%inisial
mixing proportion
p=8;%jumlah variabel
N=17;%jumlah parameter per cluster
nu=5;% inisial derajat bebas
 [~,mu,~]=kmeans(k3_0212,p); %inisialisasi rata-rata menggunakan k-
means
for j=1:p
    Sigma{j}=eye(p); %inisialisasi matriks Sigma= matiriks identitas
    ukuran pxp sebanyak p
end

[bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=RMBCmml(k3_0212,Pj,mu,Sigm
a,nu,N,5000,'agglomerate')
```

### Lampiran 31 (lanjutan)

```
Pj=[0.125,0.125,0.125,0.125,0.125,0.125,0.125,0.125];%inisial
mixing proportion
p=8;%jumlah variabel
N=17;%jumlah parameter per cluster
nu=5;% inisial derajat bebas
[~,mu,~]=kmeans(k4_0212,p); %inisialisasi rata-rata menggunakan k-
means
for j=1:p
Sigma{j}=eye(p); %inisialisasi matriks Sigma= matiriks identitas
ukuran pxp sebanyak p
end

[bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=RMBCmml(k4_0212,Pj,mu,Sigm
a,nu,N,5000,'agglomerate')

Pj=[0.3,0.3,0.4];%inisial mixing proportion
p=3;%jumlah variabel
N=7;%jumlah parameter per cluster
nu=5;% inisial derajat bebas
[~,mu,~]=kmeans(k5_0212,p); %inisialisasi rata-rata menggunakan k-
meansc
for j=1:p
Sigma{j}=eye(p); %inisialisasi matriks Sigma= matiriks identitas
ukuran pxp sebanyak p
end

[bestPj,bestmu,bestSigma,nu,indZ,Z,U,M]=RMBCmml(k5_0212,Pj,mu,Sigm
a,nu,N,5000,'agglomerate')
```

Lampiran 32. Output Kelompok Optimal RMBC-MML Pada Subset Data  $k_2$  0212,  $k_2$  0213,  $k_2$  0214, dan  $k_5$  0213

```
>> load('E:\output Matlab\k2 0212 g2 nu7.mat')
>> indZ

indZ =
    Columns 1 through 22
    2     1     1     1     1     1     1     1     2     2     2     2
       2     2     2     2     1     1     1     1
    Columns 23 through 33
    2     2     1     1     1     2     1     1     1     1     1

>> load('E:\output Matlab \k2 0213 g2 nu7.mat')
>> indZ

indZ =
    Columns 1 through 22
    2     1     1     1     1     1     1     1     2     2     2     2
       2     2     2     1     1     1     1     1
    Columns 23 through 33
    2     1     1     1     1     1     1     1     1     1     1

>> load('E:\output Matlab \k2 0214 g2 nu7.mat')
>> indZ

indZ =
    Columns 1 through 22
    1     2     2     2     2     2     2     2     2     1     1     2     1
       2     1     1     1     2     2     2     2     1     1
    Columns 23 through 33
    2     2     2     2     2     2     2     2     2     2     2

>> load('E:\output Matlab \k5 0213 g2 nu7.mat')
>> indZ

indZ =
    Columns 1 through 22
    1     1     1     1     1     1     1     1     1     1     2     2     1
       1     1     2     1     1     1     1     1     1     1
    Columns 23 through 33
    2     1     1     1     1     1     1     1     1     1     1
```

Lampiran 33. Perbandingan Keanggotaan Kelompok Optimal MBC-ICL Dengan RMBC-MML Pada Subset Data  $k_2$  0212,  $k_2$  0213,  $k_2$  0214, dan  $k_5$  0213

PROVINSI	k2 0212		k2 0213		k2 0214		k5 0213	
	ICL	MML	ICL	MML	ICL	MML	ICL	MML
ACEH	1	1	1	1	1	1	1	1
SUMATERA UTARA	1	1	1	1	1	1	1	1
SUMATERA BARAT	1	1	1	1	1	1	1	1
RIAU	1	1	1	1	1	1	2	1
JAMBI	1	1	1	1	1	1	1	1
SUMATERA SELATAN	1	1	1	1	1	1	1	1
BENGKULU	1	1	1	1	1	1	1	1
LAMPUNG	1	1	1	1	1	1	1	1
KEP. BANGKA BELITUNG	2	2	2	2	2	2	1	1
KEP. RIAU	2	2	2	2	2	2	2	2
DKI JAKARTA	2	2	1	2	1	1	2	2
JAWA BARAT	2	2	2	2	2	2	1	1
JAWA TENGAH	2	2	2	2	2	2	1	1
DI YOGYAKARTA	2	2	1	2	1	1	1	1
JAWA TIMUR	2	2	2	2	2	2	1	1
BANTEN	2	2	2	2	2	2	2	2
BALI	2	2	1	1	2	2	1	1
NUSA TENGGARA BARAT	1	1	1	1	1	1	1	1
NUSA TENGGARA TIMUR	1	1	1	1	1	1	1	1
KALIMANTAN BARAT	1	1	1	1	1	1	1	1
KALIMANTAN TENGAH	1	1	1	1	1	1	1	1
KALIMANTAN SELATAN	2	2	2	2	2	2	1	1
KALIMANTAN TIMUR	2	2	1	2	1	1	2	2
SULAWESI UTARA	2	2	1	1	1	1	1	1
SULAWESI TENGAH	1	1	1	1	1	1	1	1
SULAWESI SELATAN	1	1	1	1	1	1	1	1
SULAWESI TENGGARA	1	1	1	1	1	1	1	1
GORONTALO	2	2	1	1	1	1	1	1
SULAWESI BARAT	1	1	1	1	1	1	1	1
MALUKU	1	1	1	1	1	1	1	1
MALUKU UTARA	1	1	1	1	1	1	1	1
PAPUA BARAT	1	1	1	1	1	1	1	1
PAPUA	1	1	1	1	1	1	1	1

Lampiran 34. Hasil Pengelompokan dan Parameter Subset Data  $k_5$  0815  
Menggunakan RMBC-MML

```
>> load('k5_0815.mat')
>> indZ
indZ =
  Columns 1 through 22
      2      1      2      2      1      1      1      2      2      2      2      1
2      2      2      2      2      2      1      1      2      2
  Columns 23 through 33
      2      2      2      2      2      1      1      2      1      1      1
>> bestmu
bestmu =
    64.5166    65.9446    61.4799
    62.1227    56.0241    50.3027
>> bestPj
bestPj =
    (1,1)      0.3186
    (1,2)      0.6814
>> bestSigma{1,1}
ans =
    24.6032    36.1257    41.6155
    36.1257    75.1079    81.9773
    41.6155    81.9773    90.4163
>> bestSigma{1,2}
ans =
    18.6355     7.6901     4.0051
     7.6901   159.9064   154.8277
     4.0051   154.8277   151.2472
```

*Halaman ini sengaja dikosongkan*

## BIOGRAFI PENULIS



**Mety Agustini** lahir di Kelapa Kampit, Belitung pada tanggal 30 Agustus 1983. Penulis merupakan putri kelima dari enam bersaudara, pasangan Bapak Alm. Djamli Semaun dan Ibu Farima. Penulis menempuh pendidikan formal di SDN 12 Kelapa Kampit (1990-1996), SLTPN 4 Sungailiat (1996-1999) dan SMUN 2 Sungailiat (1999-2002). Selanjutnya, penulis melanjutkan pendidikan di Sekolah Tinggi Ilmu Statistik (STIS) Jakarta (2002-2006)

Jurusan Statistik Ekonomi. Setelah menyelesaikan pendidikan DIV di STIS, penulis ditugaskan bekerja di BPS Kabupaten Bangka Tengah, BPS Provinsi Bangka Belitung dan BPS Kabupaten Bangka. Pada tahun 2015 penulis memperoleh kesempatan untuk melanjutkan pendidikan Pasca Sarjana Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Pembaca yang ingin memberikan kritik, saran, dan pertanyaan mengenai penelitian ini dapat menghubungi penulis melalui email [mety.assahid@bps.go.id](mailto:mety.assahid@bps.go.id).